# RET receptor tyrosine kinase architecture, protein interactions and chemical inhibition

**Kerry Marie Goodman**

University College London

&

Cancer Research UK London Research Institute

PhD Supervisor: Neil McDonald

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2012

## Declaration

I Kerry Marie Goodman confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## Abstract

The RET receptor tyrosine kinase plays a major role in embryonic and adult vertebrate development. Deregulation of RET signalling leads directly to multiple human diseases. Like many receptor tyrosine kinases, the intracellular tyrosine kinase domain of RET is activated by binding of ligand/co-receptor to its ectodomain (ECD). RET is unique among receptor tyrosine kinases, possessing cadherin-like domains (CLD1–4) within its ECD. This thesis describes the architecture of the RET-ECD elucidated using small-angle X-ray scattering (SAXS). These data reveal the interdomain angles and non-linearity of the four CLDs, as well as the location of the membrane-proximal cysteine rich domain (CRD) packed against CLD4. I use this SAXS-derived model of the RET-ECD together with published crystal structures of a RET ligand and co-receptor, to fit into a negative stain electron microscopy 3D reconstruction of a mammalian ligand/co-receptor/RET-ECD ternary complex. The resulting preliminary pseudo-atomic model contains a two-fold symmetrical RET ternary complex with two RET-ECDs wrapped around a core ligand/co-receptor, making extensive contacts from both the N-terminal CLD1–3 region and the membrane-proximal CRD consistent with previous biochemical data and our antibody-epitope mapping. This thesis describes the first view of the RET-ECD and the ligand/co-receptor/RET ternary complex architecture, with important implications for Hirschsprung's disease and for understanding how ligand-independent RET activation occurs in type 2 multiple endocrine neoplasias.

# Acknowledgement

The lab in which one carries out a PhD has a huge influence on how much you learn, how much you achieve, how much you grow as a scientist and a person and how enjoyable the whole process is. Mine has been fantastic and I'm grateful to every member. Thanks to Neil, my supervisor, for his support and ceaseless enthusiasm, for opening opportunities for me to take the project further and for letting me make mistakes. Thanks to Erika, for her friendship, for keeping me going when things got tough and for her guidance and advice. Thanks to Andy, for learning SAXS with me, for teaching me and helping me with crystallography, for all the 24 hour shifts and for being very patient. Thanks to Phil, for teaching me protein production and many other experimental techniques, for keeping us all going and for rescuing the insect cells on a number of occasions. Thanks to Svend, for letting me work on his project, for the long fought battle to produce the ternary complex and for his positivity. Thanks to Maureen, Emily, Erika, Emma B. and Laura for reading through this thesis.

I'd also like to acknowledge Ed Morris's group, where I've spent a fair proportion of the last 18 months. Thanks to Fabienne, for all the time she devoted to me and to this project. And thanks to Ed, Fabienne and Paula for teaching me electron microscopy and guiding me through the whole process.

Outside of science there are many people without whom the last four years would have been so much poorer. Thanks to my housemates over the years: Laura, Sophie, Mel, Martin and Helen, for being great friends and so easy to live with, and particular thanks to Helen, Mel and Laura for being there through the last year. Thanks to Chris and Emma B. for similarly being dear friends. Thanks to Monica, without whom I don't know how I'd have made it, thanks for being there and for praying. Thanks to Heather, Emma T. and Catherine, for their friendship, encouragement, guidance and prayers. Thanks to Johan, for believing in me. Thanks to my family for loving me with all of my flaws. Thanks to mum and Graham, for being there and for being such an example of dedication, love and hard work. Thanks to dad, for his support and love. Thanks to Leila and Ross, for being fantastic and making me laugh. And thanks to Helen and Stephen, for being such a joy and for making me feel loved.

Thankyou Father, for all of the above and so much more.

# Table of contents

# Table of figures

# List of tables

# Abbreviations

| | |
|---|---|
| A-loop | activation loop |
| Abs280 | absorbance at 280 nm |
| AEBSF | 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride |
| AMPPNP | 5'-adenylyl-β,γ-imidodiphosphate |
| ATP | adenosine-5'-triphosphate |
| C-lobe | carboxy-terminal lobe |
| C-tail | carboxy-terminal tail |
| cAMP | cyclic adenosine monophosphate |
| CLD | cadherin-like domain |
| CMML | chronic myelomonocytic leukaemia |
| CRD | cysteine rich domain |
| D1 | domain 1 |
| DLS | dynamic light scattering |
| Dmax | maximum vector length |
| dmRET | *Drosophila melanogaster* RET |
| DN-cadherin | *Drosophila* N-cadherin |
| DOK | downstream of tyrosine kinase |
| DTT | dithiothreitol |
| EC | extracellular cadherin |
| ECD | ectodomain or extracellular domain |
| EDTA | ethylenediaminetetraacetic acid |
| EGF | epidermal growth factor |
| EGFR | epidermal growth factor receptor |
| EM | electron microscopy |
| Endo-F1 | endoglycosidase F1 |
| Endo-H | endoglycosidase H |
| ENS | enteric nervous system |
| ER | endoplasmic reticulum |
| ERK | extracellular signal-related kinase |
| ESRF | European synchrotron radiation facility |
| FGFR | fibroblast growth factor receptor |
| FMTC | familial medullary thyroid carcinoma |

| | |
|---|---|
| GDNF | glial cell-derived neurotrophic factor |
| GFL | GDNF family ligand |
| GFRα | GDNF family receptor alpha |
| GPI | glycosylphosphatidylinositol |
| GST | glutathione S-transferase |
| hRET | human RET |
| HEK | human embryonic kidney cells |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| Hi5 | *Trichopulsia ni* derived High Five cells |
| His | hexahistidine tag |
| HPLC | high performance liquid chromatography |
| I(0) | intensity of SAXS scattering profile extrapolated to s=0 |
| ICD | intracellular domain |
| IgG | immunoglobulin G |
| IR | insulin receptor |
| JAK | janus kinase |
| JM | juxtamembrane domain |
| JNK | c-Jun N-terminal kinase |
| KD | kinase domain |
| LDH | lactate dehydrogenase |
| MAPK | mitogen-activated protein kinase |
| MEN2 | multiple endocrine neoplasia type 2 |
| MME | monomethyl ether |
| MMT | L-Malic acid, MES, Tris |
| MOI | multiplicity of infection |
| MS | mass spectrometry |
| MSA | multi-variate statistical analysis |
| mTC | mammalian ternary complex |
| MTC | medullary thyroid carcinoma |
| MuSK | muscle specific kinase |
| MW | molecular weight |
| N-linked | asparagine linked |
| N-lobe | amino-terminal lobe |
| NAG | N-acetylglucosamine |

| | |
|---|---|
| NCAM | neural cell adhesion molecule |
| NEB | New England Biolabs |
| NGF | nerve growth factor |
| NMR | nuclear magnetic resonance |
| NSD | normalised spatial discrepancy |
| pRET-KD | phosphorylated RET kinase domain |
| P(r) | pair-distance distribution function |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PEG | polyethylene glycol |
| PFU | plaque forming unit |
| PI3K | phosphatidylinositol-3-kinase |
| PK | pyruvate kinase |
| PKC | protein kinase C |
| PLC | phospholipase C |
| PNGase F | peptide N glycosidase F |
| PTC | papillary thyroid carcinoma |
| RAS | rat sarcoma oncogene |
| RET | rearranged during transfection |
| Rg | radius of gyration |
| RMSD | root mean square deviation |
| RTK | receptor tyrosine kinase |
| SAXS | small-angle (or solution) X-ray scattering |
| SCF | stem cell factor |
| SDS | sodium dodecyl sulphate |
| SEC-MALS | size exclusion chromatography coupled with multi-angle light scattering |
| SEM | standard error of the mean |
| Sf9 | *Spodoptera frugiperda* 9 cells |
| SLS | static light scattering |
| SPA | single particle analysis |
| SS | signal sequence |
| STAT | signal transducer and activator of transcription |
| TC | ternary complex |

| | |
|---|---|
| TEV | tobacco etch virus protease |
| TLS | translation, libration and screw-rotation |
| TM | transmembrane domain |
| WT | wild type |
| zRET | zebrafish RET |
| zTC | zebrafish ternary complex |

# Chapter 1. Introduction to the RET receptor tyrosine kinase

## 1.1 Receptor tyrosine kinases structure and function

Receptor tyrosine kinases (RTKs) are cell surface receptors for many soluble and membrane-anchored polypeptide growth factors and cytokines (Gschwind *et al.*, 2004). RTKs couple extracellular signals to intracellular signalling pathways. They allow cells to respond to changes in their environment and to form complex tissues and make cell-to-cell contacts. RTKs activate intracellular signalling pathways that control cellular processes such as proliferation, differentiation and apoptosis. They also have a critical role in the development and progression of many types of cancer (Lemmon and Schlessinger, 2010). RTKs have evolved with multi-cellular organisms. The intracellular signalling pathways they link to are found in unicellular organisms such as yeast, but the RTKs only appear in evolution from the invertebrate lineages (Manning *et al.*, 2002a).

There are 58 known RTKs in the human genome, which fall into 20 subfamilies. Well known examples include the epidermal growth factor receptor (EGFR) and the insulin receptor. The majority of RTKs are single pass transmembrane receptors and have a common overall architecture, consisting of a highly variable ligand-binding ectodomain (ECD), a transmembrane domain (TM), and an intracellular domain (ICD) containing a highly conserved tyrosine kinase that is frequently flanked by the more divergent juxtamembrane domain (JM) and C-terminal tail (C-tail) (Lemmon and Schlessinger, 2010).

RTK-ECDs are diverse in both structure and sequence, but are enriched for globular protein interaction modules such as immunoglobulin C2-like domains and fibronectin type III domains (Lemmon and Schlessinger, 2010). These domains can be organised repetitively or in different combinations to generate ligand-binding sites for their cognate ligands (examples of the diverse RTK ECD architecture are shown in Figure 1.1). How RTK-ECDs couple with and stimulate the catalytic activity of associated intracellular tyrosine kinase function is a crucial question and likely to be distinct for

each type of RTK as the ligands are quite unrelated in many cases. Commonalities for RTK activation and auto-inhibition are however expected and are indeed observed. For example: control of RTK oligomeric state, allosteric activation and the removal of auto-inhibitory elements in response to ligand-binding (Hubbard and Miller, 2007).

The ICDs of RTKS are more structurally similar, possessing tyrosine kinase domains (KDs), which share their overall fold with all tyrosine and serine/threonine kinases (Cowan-Jacob, 2006). The KDs are flanked by a JM and C-tail, which are commonly flexible and are important for kinase regulation and downstream signalling.

RTKs respond to their cognate ligands by activation of their intracellular kinase activity, which is commonly associated with dimerisation of the receptor (Hubbard and Miller, 2007). This results in autophosphorylation *in trans* of a number of tyrosine residues within RTK ICDs, with phosphorylation of the different sites occurring in a precise sequential order (Furdui *et al.*, 2006). These phosphotyrosines contribute to increased kinase activity and/or act as recruitment sites for signalling molecules containing SH2 (Src homology-2) and phosphotyrosine-binding (PTB) domains. Recruitment of these signalling molecules results in the activation of interconnected downstream signalling pathways, leading to a coordinated response by a diverse range of cellular processes, involving for example changes in gene expression and cellular metabolism (Lemmon and Schlessinger, 2010).

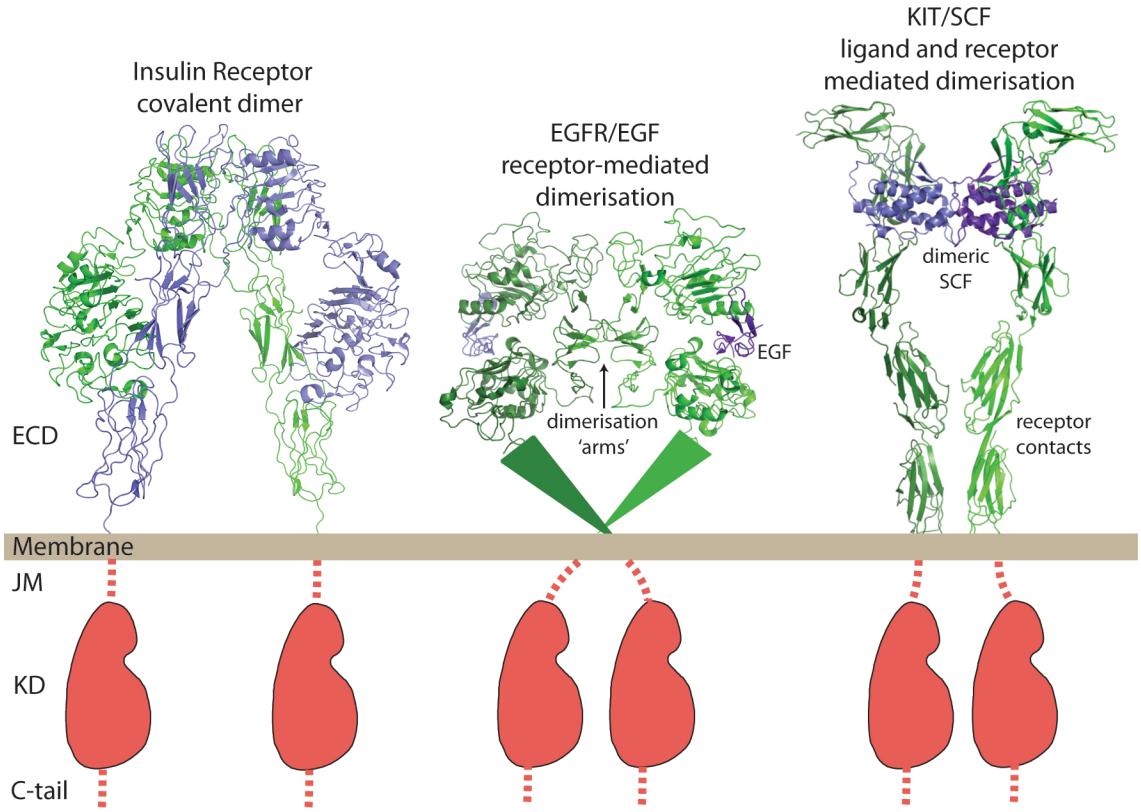



**Figure 1.1 RTK architecture and dimerisation mechanisms**

Ribbon representation of known crystal structures of RTK-ECDs in their ligand-bound dimeric form. Three different RTK superfamily members are highlighted, indicating the structural diversity of the ligand-binding extracellular regions of RTKs and some of the distinct mechanisms of RTK dimerisation that have been observed. The highly similar intracellular regions are shown as schematic lollipops for simplicity. Dotted lines represent the commonly flexible JMs and C-tails. The insulin receptor ECD is a disulphide-linked dimer of αβ heterodimers, consisting of three fibronectin type III domains, two leucine-rich domains and a cysteine rich domain. One αβ heterodimer is coloured slate and the other green. The epidermal growth factor receptor (EGFR)-ECD dimer (green and dark green) bound to its monomeric ligand EGF (slate and purple), exemplifies ligand-induced receptor-mediated dimerisation (Lemmon, 2009). Each EGFR monomer consists of two leucine-rich domains and two cysteine rich domains. The second cysteine rich domain is not observed in the crystal structure and is represented schematically as a triangle. The KIT-ECD/stem cell factor (SCF) complex exemplifies dimeric-ligand-mediated RTK dimerisation. The KIT-ECDs each consist of five immunoglobulin domains and are coloured green and dark green. The SCF monomers are coloured slate and purple. (Crystal structure references: insulin receptor PDB 2DTG (McKern *et al.*, 2006); EGFR/EGF PDB 1IVO (Ogiso *et al.*, 2002); KIT/SCF PDB 2E9W (Yuzawa *et al.*, 2007)).

### 1.1.1 Protein kinase structure and active conformation

Tyrosine kinases are phosphoryl transfer reaction catalysts. Upon binding of ATP and a tyrosine-containing substrate to a catalytically productive conformation of the kinase, they mediate the transfer of the ATP γ phosphate to the hydroxyl group of a substrate tyrosine residue (Figure 1.2a). The tyrosine kinase catalytic base (D1132 in the insulin receptor kinase) deprotonates the substrate tyrosine, catalysing nucleophilic attack by the tyrosine oxygen onto the electrophilic γ phosphate phosphorous. The reaction proceeds through a pentavalent transition state, resulting in a phosphorylated tyrosine and ADP. These products then dissociate from the kinase allowing a further ATP and substrate to bind and a subsequent round of catalysis to proceed (Lassila *et al.*, 2011).

The RTK-KDs like other tyrosine (and serine-threonine) kinases possess a bilobal fold, with a smaller predominantly β-strand N-lobe and a larger α-helical C-lobe (Figure 1.2) (Cowan-Jacob, 2006). The strict conformational requirements within the RTK-KD fold for a catalytically productive kinase are shared with most other eukaryotic kinases. The structural elements crucial for catalytic activity are illustrated in Figure 1.2. The requirements for catalysis include: (1) the αC helix positioned close to the catalytic centre, which allows formation of the structurally and catalytically important K1030-E1074 salt bridge (insulin receptor numbering); (2) an open and ordered activation loop conformation, allowing substrate binding; (3) the magnesium-binding *DFG* motif in the '*DFG-in*' arrangement, allowing proper magnesium coordination by the *DFG* aspartate; (4) the correct relative position of the N and C-lobes, bringing catalytic residues from both lobes together (Huse and Kuriyan, 2002). Disruption of the active arrangement of any of these elements will result in the inhibition of protein kinase activity.

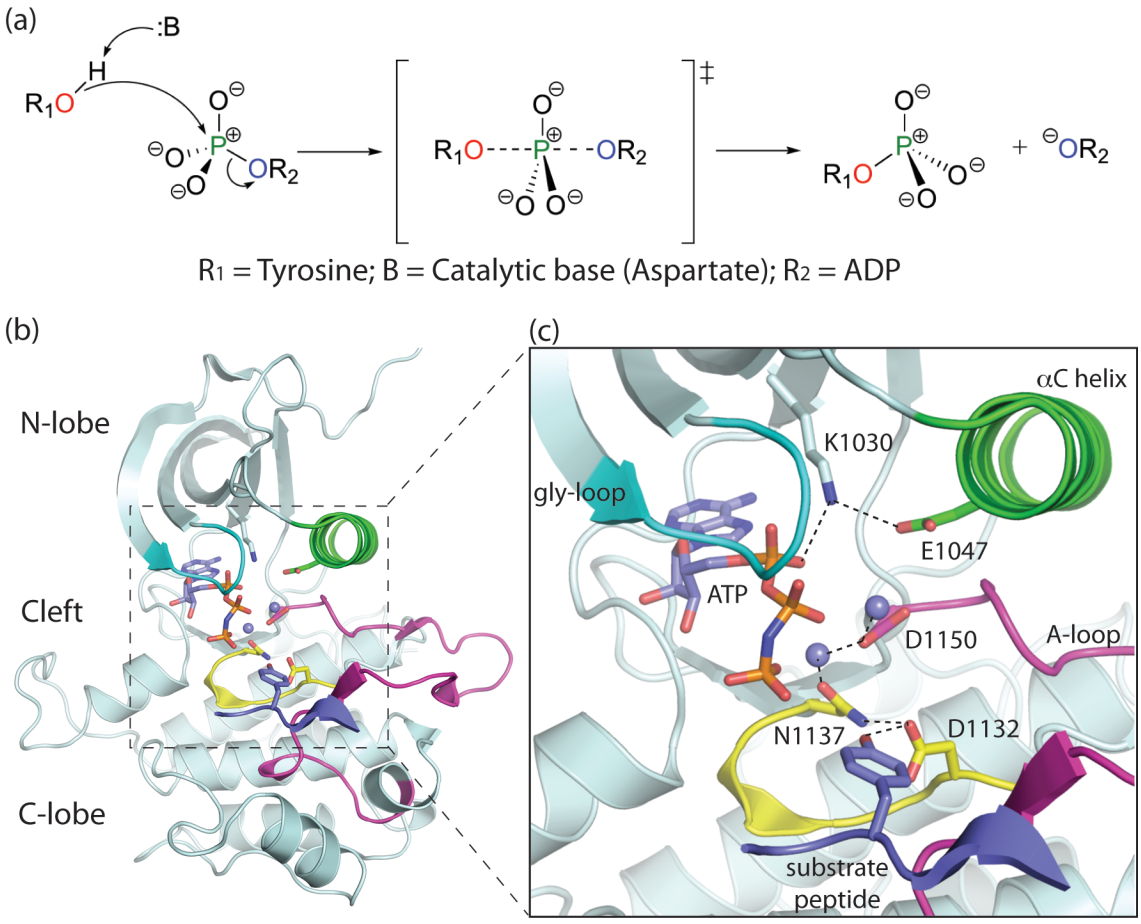



**Figure 1.2 Protein tyrosine kinase domain active conformation**

(a) Catalytic mechanism proposed for the phosphoryl transfer reaction of a tyrosine kinase. (b) & (c) The crystal structure of the insulin receptor KD (pale blue ribbon), with an ATP-analogue (purple sticks), substrate peptide (purple ribbon) and two magnesium ions (purple spheres) bound (PDB 1IR3; Hubbard, 1997). The key structural elements in an active kinase conformation are coloured as follows: αC helix, green; glycine-rich loop, cyan; catalytic loop, yellow; activation loop including the substrate-binding P+1 loop, pink. Side chains of conserved residues important for catalysis are shown as sticks and dashed black lines highlight key interactions. D1132 is the catalytic base, which interacts with the attacking nucleophilic hydroxyl on the substrate peptide tyrosine. N1137 positions D1132 for catalysis and is involved in magnesium coordination. Other key residues are discussed in the text.

## 1.2 Regulation of RTK activity

In the simplest view of RTK function, their role in growth factor signalling pathways is as a growth factor responsive switch. In the absence of growth factor their intracellular tyrosine kinase activity is inhibited, then upon binding of the growth factor to their ECD the kinase is activated in a reversible manner. Structural biology has played a key role

in understanding how different RTKs ensure that kinase activity is linked to growth factor binding, although no RTK is fully understood. The precise details of regulatory mechanisms differ between individual RTKs, although there are a number of common themes that have emerged from the RTKs studied to date. These themes include allostery, dimerisation and autophosphorylation *in trans* (Hubbard and Miller, 2007).

### 1.2.1 Allostery and RTK activation

Within a broad definition of allostery, as adopted by Kuriyan and Eisenberg (Kuriyan and Eisenberg, 2007), allosteric regulation may become a common theme in understanding RTK regulation (Jura *et al.*, 2011). They define allostery as where the activity of a protein is regulated by communication between the active site and a distant site of modification or binding. Upon modification/binding at this allosteric site, a structural change affecting tertiary and/or quaternary structure or a change in the flexibility of the protein occurs, leading to modulation of the activity of the protein at the active site (Kuriyan and Eisenberg, 2007).

EGFR has emerged as the archetypal example of this type of regulation in a RTK. Autophosphorylation of EGFR does not result in kinase activation, rather an allosteric mechanism, similar to that observed in activation of CDKs by cyclins activates EGFR kinase activity following ligand-induced dimerisation (Zhang *et al.*, 2006). This involves the formation of an asymmetric dimer of the EGFR-KDs, where one KD allosterically activates the other KD by binding and altering the position of its αC helix. This asymmetric KD dimer is stabilised by interactions between the two JMs (Jura *et al.*, 2009), which may also be coupled to a particular arrangement of the TMs (Bocharov *et al.*, 2008; Endres *et al.*, 2011).

### 1.2.2 The role of dimerisation in RTK activation

RTKs are activated in ligand/receptor complexes containing at least a receptor dimer, if not a higher oligomer (Heldin, 1995; Ullrich and Schlessinger, 1990). The precise role of ligand-driven dimerisation in the activation of RTKs varies amongst the superfamily (Lemmon and Schlessinger, 2010). For example, some receptors adopt pre-formed dimers in the absence of ligands and undergo allosteric changes in response to ligand rather than a change in receptor oligomeric state. Dimerisation appears to be a pre-

requisite for RTK activation, facilitating both allosteric and *trans*-autophosphorylation mechanisms of kinase activation (Chen *et al.*, 2008; Zhang *et al.*, 2006). Furthermore activation of downstream signalling pathways is dependent on *trans*-autophosphorylation of tyrosine residues within the RTK ICD requiring close receptor proximity to allow phosphorylation *in trans* to occur. The method of RTK dimerisation adopted by individual RTKs varies amongst the superfamily as described below.

#### *1.2.2.1 Ligand-independent RTK dimerisation*

At one end of the spectrum the insulin receptor family members are constitutive disulphide-linked dimers, with dimerisation therefore not part of the activation mechanism (McKern *et al.*, 2006). Ligand-induced non-covalent dimerisation of RTKs is proposed to be a more widespread feature of RTK regulation (Lemmon and Schlessinger, 2010), although evidence is accumulating for pre-formed dimers of some RTKs in the absence of ligand (Whitty and Riera, 2008), as observed in microscopy studies of EGFR, where ligand then appears to induce tetramerisation (Clayton *et al.*, 2005; Kastner *et al.*, 2009; Webb *et al.*, 2008).

Where ligand-independent dimerisation occurs, whether constitutive or transient, auto-inhibitory mechanisms are needed to prevent stochastic misfiring of the kinase in the dimer in the absence of ligand. In the case of the constitutively dimeric insulin receptor, binding of a dimeric pseudosubstrate inhibitory partner protein to both KDs in the receptor dimer inhibits kinase activity (Depetris *et al.*, 2005). Activation of the kinase activity of RTK's in pre-formed dimers/oligomers is allosteric, with the ligand inducing a conformation change in the RTK dimer, which leads to kinase activation (Hubbard and Miller, 2007). In the case of the insulin receptor this would include the dissociation of the inhibitory partner protein.

#### *1.2.2.2 Ligand-dependent RTK dimerisation*

Observations from many RTKs suggest that ligand-induced RTK dimerisation is a widespread mode of RTK dimerisation but with many variations on the theme (Lemmon and Schlessinger, 2010). Four examples of these variations include: (1) dimeric-ligand-mediated dimerisation, where the dimeric-ligand exclusively mediates receptor dimerisation (e.g. TrkA/NGF (nerve growth factor); Wehrman *et al.*, 2007); (2)

dimeric-ligand and receptor-mediated dimerisation, where the dimeric-ligand along with direct receptor:receptor interactions mediate dimerisation (e.g. KIT/stem cell factor (SCF); Yuzawa *et al.*, 2007); (3) monomeric-ligand and receptor-mediated dimerisation, where the two monomeric ligands both contact both receptors in the dimer and receptor:receptor contacts are also involved (e.g. FGFR (fibroblast growth factor receptor)/FGF; Schlessinger *et al.*, 2000); (4) receptor-mediated dimerisation, where the dimer is exclusively formed from receptor:receptor contacts (e.g. EGFR/EGF; Ogiso *et al.*, 2002). Even where dimerisation is not ligand dependent, the ligand still binds and stabilises a particular kinase-active dimeric/oligomeric receptor arrangement and therefore these ligand-engagement mechanisms are still relevant (Lemmon and Schlessinger, 2010).

Where receptor dimerisation is a key part of the RTK activation process, inhibitory mechanisms need to be in place to prevent 'active' dimer formation in the absence of ligand. Such mechanisms would then be overridden upon ligand binding. A structurally characterised example of the class 2 mode of dimerisation, described above, is KIT, a member of the PDGFR family. The KIT-ECD consists of five immunoglobulin (Ig)-like domains, which are arranged end-to-end creating an overall quasi-linear architecture. SCF, its dimeric-ligand, binds KIT's three N-terminal membrane-distal Ig-like domains, cross-linking two receptors (Figure 1.1). This binding and dimerisation event leads to a rearrangement of the two membrane-proximal Ig-like domains, resulting in receptor:receptor interactions between these domains near the cell surface. It is thought that charge repulsion prevents these low-affinity receptor:receptor interactions occurring in the absence of ligand. This is overcome by ligand binding, which increases the local concentration of these two domains and positions them poised for productive interaction. It is the interaction of these membrane-proximal domains that is thought to lead to kinase activation, since it brings the TMs together, and therefore likely the tyrosine kinase domains (Yuzawa *et al.*, 2007).

In the case of EGFR (and family members ErbB3 and ErbB4), prevention of ligand-independent 'active' dimer formation is achieved by the inaccessibility of the EGFR 'dimerisation arm' in the absence of the EGF ligand binding. EGFR dimerisation upon EGF binding follows an extensive ligand-induced conformational change within the EGFR-ECD, which exposes the previously buried 'dimerisation arm' (Bouyain *et al.*,

2005; Cho and Leahy, 2002; Ferguson *et al.*, 2003; Ogiso *et al.*, 2002) (Figure 1.1). Alongside the global rearrangement of the EGFR-ECD, EGF binding also results in the bending of the domain primarily involved in dimerisation (domain 2) (Alvarado *et al.*, 2009). This bend is required in at least one of the EGFR molecules in the dimer for kinase activation, since it allows the receptor dimer to adopt a conformation that positions the TMs (and therefore the ICDs) productively (Liu *et al.*, 2012). This is thought to be the basis for the regulation of EGFR family member ErbB2, which has no known ligand and has a constitutively exposed dimerisation arm (Garrett *et al.*, 2003). ErbB2 does not have a bent domain 2 and therefore ErbB2 homodimers are unable to form an active dimeric conformation (Alvarado *et al.*, 2009). They must rather heterodimerise with a ligand-bound member of the EGFR family (Liu *et al.*, 2012).

### 1.2.3 Regulation by autophosphorylation

Protein structures are dynamic and the tyrosine kinase domain is likely to sample a range of conformations in solution. In the low activity (auto-inhibited) form of tyrosine kinase domains, a broad range of mostly inactive conformers are thought to be sampled (Chen *et al.*, 2007). Upon phosphorylation, flexible regions with phosphorylation sites become tethered to phospho-binding pockets restricting the range of conformers in solution towards the single active conformation (Chen *et al.*, 2007). Static inhibitory conformations have been defined for some tyrosine kinases, such as the insulin receptor described below, but for others there is little evidence for discrete inhibited forms. Autophosphorylation follows ligand engagement, which is thought to arrange the KDs in an RTK dimer/oligomer such that autophosphorylation *in trans* can occur despite low basal kinase activity (Furdui *et al.*, 2006).

There are three elements within the ICDs of RTKs that can regulate kinase activity in this phosphorylation dependent way: the KD activation loop, the JM and the C-tail (Hubbard and Miller, 2007; Lemmon and Schlessinger, 2010). These are discussed below.

#### *1.2.3.1 Activation loop phosphorylation*

The activation loop of protein kinases lines the base of the kinase cleft, between the N and C-lobes. It is located between C-lobe helices αE and αF and is generally 20–30

amino acids long. The activation loop begins with the conserved DFG motif and ends with a conserved APE motif. The central region is less well conserved between kinases, although tyrosine kinases commonly have between 1 and 3 tyrosines, which are autophosphorylation sites (Nolen *et al.*, 2004).

Crystal structures of non-phosphorylated KDs commonly display activation loop conformations that are incompatible with kinase activity (e.g. EphB2 (Wiesner *et al.*, 2006); FGFR (Chen *et al.*, 2007); insulin receptor (Hubbard *et al.*, 1994)). The non-phosphorylated insulin receptor KD structure shows the activation loop blocking both the ATP binding pocket and the active site (Figure 1.3) (Hubbard *et al.*, 1994). This non-productive activation loop arrangement is also linked to the non-productive position of the αC helix, which is rotated away from the active site. This arrangement is stabilised by activation loop Y1162 binding in the active site as if it were a substrate. However since the ATP pocket is blocked, and the αC helix is improperly positioned, Y1162 is not phosphorylated *in cis* (Hubbard *et al.*, 1994).

The insulin receptor activation loop contains three tyrosine residues that are autophosphorylation sites and phosphorylation increases kinase activity 200×. The crystal structure of the insulin receptor KD phosphorylated on all three of these sites and in complex with an ATP-analogue and substrate peptide, revealed how phosphorylation of these activation loop tyrosines stabilise the kinase in an active conformation (Hubbard, 1997). Phosphorylation of Y1162 inhibits binding in its own active site. Phosphorylation of Y1163 stabilises the active, open, conformation of the activation loop by forming hydrogen bonds with two other residues in the activation loop, R1155 and G1166. Phosphorylation of Y1158 does not appear to be important for the activation loop conformation: Its role is probably simply as a recruitment site for downstream signalling molecules (Hubbard, 1997).

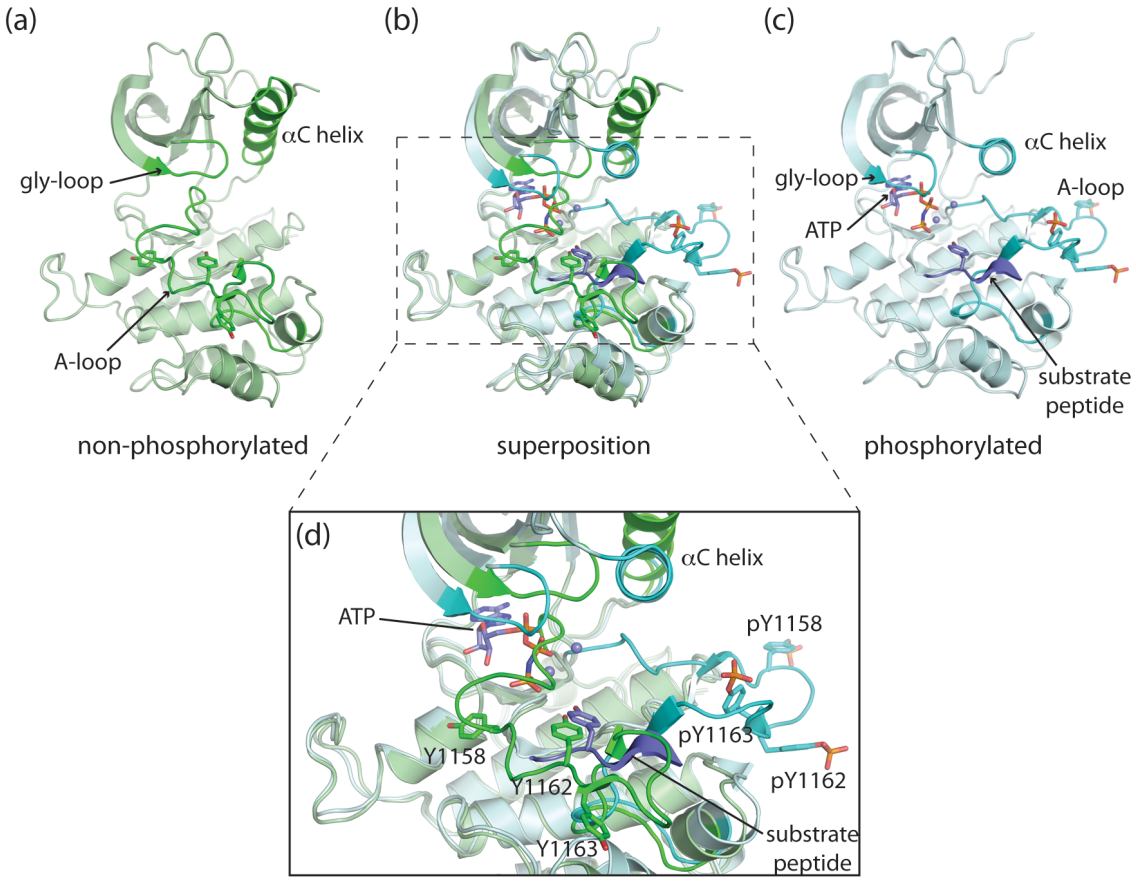



**Figure 1.3 Activation loop conformations of non-phosphorylated and phosphorylated insulin receptor KD**

(a) The crystal structure of the non-phosphorylated insulin receptor KD (pale green ribbon; PDB 1IRK; Hubbard *et al.*, 1994), with the glycine-rich loop, αC helix and activation loop highlighted (green). The side chains of the activation loop tyrosines are shown as sticks. (c) The crystal structure of the tri-phosphorylated insulin receptor KD (pale blue ribbon), with an ATP-analogue (purple sticks), substrate peptide (purple ribbon) and two magnesium ions (purple spheres) bound (as Figure 1.2; PDB 1IR3; Hubbard, 1997). The glycine-rich loop, αC helix and activation loop are highlighted (cyan) and the side chains of the phosphorylated tyrosines are shown as sticks. (b) Superposition of the non-phosphorylated and tri-phosphorylated insulin receptor KD structures. (d) Close up of the kinase cleft.

#### *1.2.3.2 Juxtamembrane domain and C-terminal tail inhibition and phosphorylation*

Often alongside regulation of kinase activity by activation loop phosphorylation, the commonly flexible JM and C-tail of RTKs can also regulate kinase activity in a phosphorylation dependent manner. For example the JM of the EphB2 receptor binds the KD *in cis* and stabilises the αC helix in a distorted non-productive conformation.

Dissociation of the JM allows the αC helix to adopt an active conformation, and JM dissociation is promoted by phosphorylation of two JM tyrosine residues (Wiesner *et al.*, 2006; Wybenga-Groot *et al.*, 2001). Tie2 is an example of C-tail inhibition. Its C-tail binds to the KD blocking substrate binding, which is likely relieved upon phosphorylation of tyrosines within the C-tail (Niu *et al.*, 2002).

## 1.3 The RET receptor tyrosine kinase

RET (rearranged during transfection) was originally discovered as a proto-oncogene that could be activated by DNA rearrangement. This rearrangement resulted in the production of a fusion protein, which could transform NIH-3T3 cells (Takahashi *et al.*, 1985). The C-terminal portion of this gene fusion product was later shown to express a transmembrane and a tyrosine kinase domain (Takahashi and Cooper, 1987). This portion came from a gene encoding a receptor tyrosine kinase (RTK), which inherited the name RET.

RET is the sole member of its RTK subfamily, and uniquely contains cadherin-like domains (CLDs) in its ECD (Anders *et al.*, 2001), as well as a cysteine rich domain (CRD). Whilst the RET-ECD is unique, the RET-KD is highly similar, in sequence and structure, to other RTK-KDs.

The intracellular region of RET is most closely related by sequence homology, to the FGFRs, vascular endothelial growth factor receptors (VEGFRs) and PDGFs (Manning *et al.*, 2002b). These RTKs are all dubbed ‘split-kinase’ RTKs since their KDs contain a hydrophilic ‘kinase insert’, an embellishment on the common KD fold and a potential source of further regulation (Ullrich and Schlessinger, 1990). There are three alternatively spliced forms of RET, which differ in their C-terminal sequences. They are called RET9, RET43 and RET51 based on the number of splice variant specific residues found at their C-termini (9, 43 and 51 respectively) (Plaza-Menacho *et al.*, 2006).

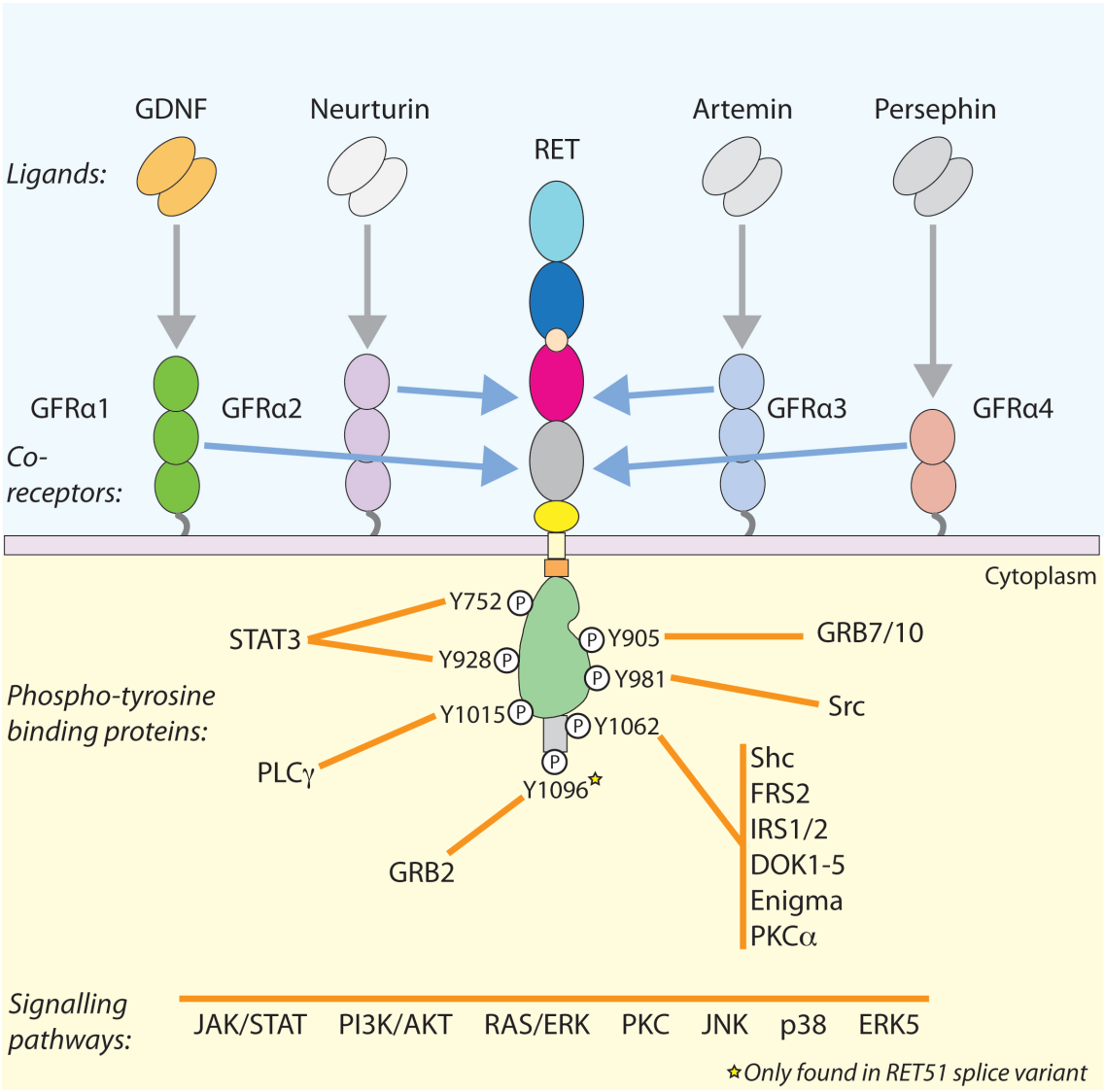



**Figure 1.4 RET canonical signalling pathway**

Cartoon schematic showing the canonical RET signalling pathway. The disulphide-linked homodimeric GDNF family ligands, with their respective GPI-linked GFRα co-receptors, bind RET and activate its intracellular tyrosine kinase activity. This leads to autophosphorylation on a number of phosphotyrosine sites within the RET-ICD, as depicted (Kawamoto *et al.*, 2004; Salvatore *et al.*, 2000). These serve as docking sites for intracellular signalling/scaffold proteins. Recruitment of these proteins to RET leads to the activation of downstream intracellular signalling pathways (Ichihara *et al.*, 2004; Kodama *et al.*, 2005; Morandi *et al.*, 2011; Plaza-Menacho *et al.*, 2006; Santoro *et al.*, 2004). DOK = downstream of tyrosine kinase; FRS = fibroblast growth factor receptor substrate 2; GRB = growth factor receptor bound protein; IRS = insulin receptor substrate; PLC = phospholipase C (other acronyms defined in text).

### 1.3.1 RET ligands and co-receptors

RET remained an orphan receptor for many years, lacking a known activating ligand. In 1996 it was shown to be activated by and mediate downstream signalling from GDNF (glial cell line derived neurotrophic factor), a distant member of the TGF-β superfamily (Durbec *et al.*, 1996; Trupp *et al.*, 1996). RET is now known to act as a receptor for the four GDNF family ligands (GFLs): GDNF, Neurturin, Artemin and Persephin; which bind RET in complex with a ligand-specific co-receptor: GFRα1, 2, 3 and 4 (GDNF family receptor α 1–4) (Figure 1.4) (Airaksinen and Saarma, 2002; Klein *et al.*, 1997; Treanor *et al.*, 1996). Only one other RTK requires a co-receptor for ligand-binding: MuSK (muscle specific kinase) responds to its ligand, agrin, via agrin binding to Lrp4 (LDL receptor-related protein 4) (Kim *et al.*, 2008; Zhang *et al.*, 2008). However there are a number of differences between the ligand/co-receptors for MuSK and RET, and therefore the mechanism of receptor activation is likely very different (Parkash *et al.*, 2008; Wang *et al.*, 2006; Zong *et al.*, 2012).

The current model for an active RET signalling complex consists of a disulphide-linked ligand homodimer, two membrane-associated GPI (glycosylphosphatidylinositol)-linked co-receptors and two RET molecules. Co-receptor can stably bind ligand in the absence of RET and work carried out on the RET/Artemin/GFRα3 complex determined the order of assembly to be: 1 GFRα3 + Artemin dimer + 1 RET + 1 GFRα3 + 1 RET (Schlee *et al.*, 2006).

However this is apparently not the only RET signalling competent complex. Soluble GFRα4, a splice variant of GFRα4, which lacks the GPI-link and contains an alternative second (D3) domain, has been shown to bind and activate RET in the absence of its ligand Persephin. This may be due to the formation of disulphide-linked homodimers of soluble GFRα4, utilising cysteine residues from the alternative D3 domain, thereby negating the need for ligand mediated dimerisation of the co-receptor (Yang *et al.*, 2007). A further alternative RET signalling complex has recently been proposed, linking RET to a new set of ligand/co-receptors. Ephrin-A has been proposed to be a novel co-receptor for RET, with RET binding GPI-linked Ephrin-A, in competition with GFRα1, and mediating reverse Eph-A receptor:Ephrin-A signalling in motor axons (Bonanomi *et al.*, 2012). RET can also be activated through crosstalk with

other growth factor signalling pathways. The long isoform of RET (RET51) is activated in response to nerve growth factor (NGF) in sympathetic neurons through an apparently indirect, GFL/GFRα independent, mechanism (Arighi *et al.*, 2005).

Within the canonical GFL/GFRα RET signalling the location of the co-receptor can vary, leading to three models of GFL/GFRα RET signalling (Ibanez, 2010). The first of these is where GFRα and RET are co-expressed in the same cell. GFRα is commonly located in lipid rafts due to its GPI-link. Lipid rafts are signalling compartments within the cell membrane, characterised by high-levels of cholesterol and sphingolipids, and allow compartmentalisation of signalling molecules associated with the cell membrane (Hancock, 2006; Lingwood and Simons, 2010). Upon GFL binding, GFRα recruits RET to these rafts leading to RET activation *in cis* (Sariola and Saarma, 2003). A second model of GFL/GFRα RET signalling is where GFRα and RET are expressed by adjacent cells. In this scenario it has been shown that the GDNF/GFRα1 complex can bind and activate RET *in trans* (Paratcha *et al.*, 2001). Finally GFRα1 is constitutively released from the cell membrane by membrane-associated phospholipases. The resulting soluble GFRα1 can still bind GDNF and subsequently bind and activate RET, removing the constraint of the requisite coincidence of GFRα and RET expression in the same cell or adjacent cells for GPI-linked GFRα (Ibanez, 2010; Paratcha *et al.*, 2001).

#### *1.3.1.1 GFL's structure and function*

In the canonical RET signalling pathway, RET is activated upon binding to a GFL in complex with a GFRα co-receptor. There are four GFLs, which are conserved amongst vertebrates: GDNF, Neurturin, Artemin and Persephin (Hatinen *et al.*, 2007). These have distinct but overlapping functions. For example GDNF is the principal GFL involved in early enteric nervous system (ENS) development, whereas Neurturin is required for the later development and maintenance of enteric and parasympathetic neurons (Airaksinen *et al.*, 2006). Curiously Neurturin has been lost during evolution from the frog (*Xenopus tropicalis*) genome and Persephin from the chicken (*Gallus gallus*) genome, with other GFLs apparently binding GFRα2 and GFRα4 in these species. The zebrafish (*Danio rerio*) genome, along with the genomes of other fish,

encodes two GDNF and two Neurturin orthologs (Hatinen *et al.*, 2007). Despite the presence of a RET ortholog in the arthropod genome no GFL ortholog has been identified in the *Drosophila melanogaster* genome (Hatinen *et al.*, 2007).

GFLs, with their GFRα co-receptors, signal through RET and also a non-RTK cell-surface receptor, neural cell adhesion molecule (NCAM) (Paratcha *et al.*, 2003). More recently GFLs GDNF, Neurturin and Artemin, have been shown to signal through the heparin sulphate proteoglycan Syndecan-3 in GABAnergic cortical neurons. This requires the GFLs to be immobilised on the extracellular matrix and is a RET and NCAM-independent signalling pathway (Bespalov *et al.*, 2011).

GFLs are distant relatives of TGF-β and are disulphide-linked homodimers. The crystal structures of GDNF and Artemin have been solved revealing each monomer has a cystine knot fold, with two β-sheet 'fingers' at one end and an α-helix 'heel' at the other (Eigenbrot and Gerber, 1997; Silvian *et al.*, 2006). Formation of the dimer results in a two-fold symmetric structure with overlapping 'heels' and projecting 'fingers' (Figure 1.5).

#### *1.3.1.2 GFRα's structure and function*

The GFLs signal through RET via interaction with GPI-linked GFRαs. Each GFL signals through a specific GFRα—GDNF/GFRα1, Neurturin/GFRα2, Artemin/GFRα3, Persephin/GFRα4 (Figure 1.4)—although there is some crosstalk, with GDNF able to bind and signal through GFRα2 for example (Airaksinen *et al.*, 1999). The GFRαs are conserved amongst vertebrates, like the GFLs, and the fish genomes encode duplicates of some of the GFRαs with the zebrafish genome containing two *GFRα1* and two *GFRα2* genes (Hatinen *et al.*, 2007).

GFRαs contain three related domains designated D1–D3 (domain 1–domain 3), except GFRα4, which only has D2–D3 (Hatinen *et al.*, 2007). The crystal structure of D2 and D3 from GFRα1 and GFRα3 revealed each domain possesses a triangular five-helix bundle fold, constrained by five disulphide bonds, which are formed by cysteine residues conserved amongst all GFRαs (Figure 1.5) (Leppanen *et al.*, 2004; Parkash *et al.*, 2008; Wang *et al.*, 2006). D2 and D3 form a compact module, with a conserved

hydrophobic interface between the two domains (Wang *et al.*, 2006). Consistent with the absence of D1 from GFRα4, GFRα-D2D3 contains the primary GFL and RET binding sites (Wang *et al.*, 2006). D1 contains eight of the ten cysteines present in D2 and D3, and is predicted to share the same overall helical fold (Leppanen *et al.*, 2004). D1 is not required for GFL and RET binding (Scott and Ibanez, 2001), but does stabilise the complex and modulate RET signalling although it remains unclear how (Virtanen *et al.*, 2005).

Alongside the GFRαs there are some other distant members of the GFR superfamily, which contain domains sharing the cysteine pattern of GFRα D1–D3. These are GFAL (GFRα-like), GFRL (GFR-like) and Gas1 (growth arrest specific 1). They are of particular interest since, unlike the GFRαs, putative Gas1 and GFRL orthologs have been identified in arthropods as well as vertebrates, and therefore could hold the answer to the enigma of arthropod RET's ligand. Gas1, GFAL and GFRL are orphan receptors however, appearing to be unable to bind GFLs (Hatinen *et al.*, 2007). Gas1 has been reported to bind RET and has been shown to influence RET signalling, but as a negative regulator (Cabrera *et al.*, 2006; Lopez-Ramirez *et al.*, 2008).

#### *1.3.1.3 GFL/GFRα complex structure and putative RET binding site*

The crystal structures of GFL/GFRα complexes GDNF/GFRα1-D2D3 and Artemin/GFRα3-D2D3 revealed a two fold symmetric structure with a GFRα-D2D3 bound to the finger region of the GFL at either end of the GFL dimer (Figure 1.5) (Parkash *et al.*, 2008; Wang *et al.*, 2006). The GFL finger inserts into the centre of the D2 triangular helical bundle, with D3 on the outside of the complex apparently stabilising D2 (Wang *et al.*, 2006). A putative RET binding site on the GFRα-D2D3 surface was identified by analysis of conservation amongst GFRαs of surface exposed residues (Wang *et al.*, 2006). Mutation of nine of these conserved residues on GFRα1, either individually or in pairs, led to a three-fold or better reduction in RET activation, measured by RET autophosphorylation. These residues were R190, K194, R197, Q198 and K202 from D2 and R257, E323 and D324 from D3, and create a highly charged putative RET interface spanning D2 and D3 (Figure 1.5a) (Parkash *et al.*, 2008).

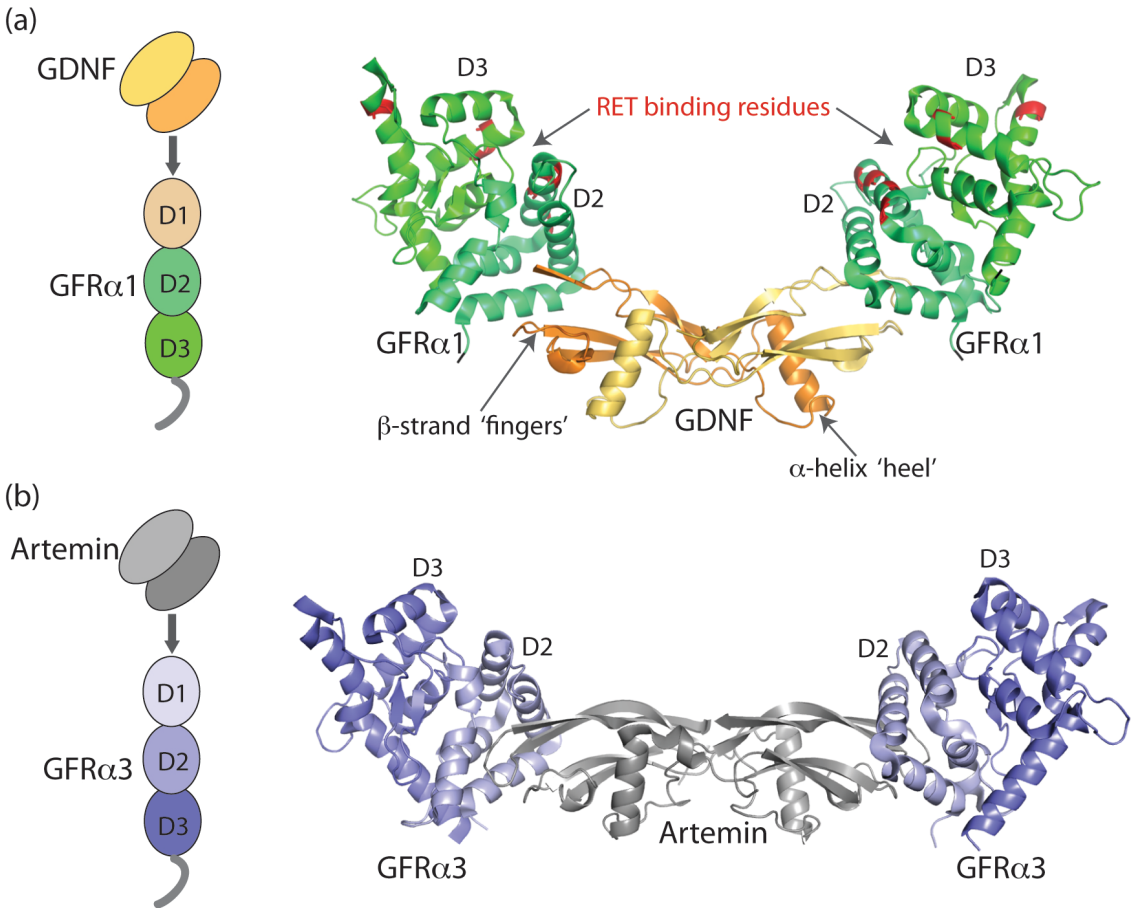



**Figure 1.5 Ligand/co-receptor crystal structures**

(a) Crystal structure of the GDNF/GFRα1-D2D3 complex shown in ribbon representation, with residues involved in RET binding highlighted in red (Parkash *et al.*, 2008). (b) Crystal structure of the Artemin/GFRα3-D2D3 complex shown in ribbon representation (Wang *et al.*, 2006). Comparison between the two structures reveals the Artemin complex is more elongated than the GDNF complex, due to the different intermonomer bend angle in the Artemin and GDNF dimers.

Comparison between the GDNF/GFRα1-D2D3 and Artemin/GFRα3-D2D3 crystal structures revealed a possible structural basis for the ability of these ligands to stimulate different downstream signalling pathways, despite signalling through the same receptor, namely RET (Parkash *et al.*, 2008). The binding of GFRα to the fingers of the GFL makes the overall shape of the complex sensitive to the bend angle of the GFL dimer. The Artemin dimer is more linear than the GDNF dimer in all of the crystal structures of both GFLs (Parkash and Goldman, 2009). This means that the Artemin/GFRα3-D2D3 complex is more linear and therefore longer than the GDNF/GFRα1-D2D3 complex (Figure 1.5), placing the putative RET binding sites on each GFRα in the

Artemin/GFRα3-D2D3 complex further apart. The arrangement of the two RET molecules in the GFL/GFRα/RET complex might therefore differ dependent on the ligand, which may lead to a different conformation of RET in its ICD as well. This could hypothetically influence downstream signalling, perhaps through accessibility of particular phosphotyrosine sites on RET (Parkash and Goldman, 2009; Parkash *et al.*, 2008).

### 1.3.2 RET's activation of intracellular signalling pathways

RET stimulates the activation of similar downstream pathways to many other RTKs, but due to its unique ligand/co-receptor, unique timing of ligand/co-receptor availability, potentially unique activation kinetics, cell type specificity and combined activation with other RTKs, cells and organs show a physiologically unique response to RET activation. The best defined hallmark of RET activation is the phosphorylation of specific and invariant tyrosine residues within the RET-ICD. Mass spectrometry analysis identified eight (Y806, Y809, Y900, Y905, Y981, Y1062, Y1090, 1096) tyrosine residues in the RET-ICD, which were phosphorylated in response to ligand/co-receptor binding (Kawamoto *et al.*, 2004). Additionally Y687, Y752, Y928, Y1015, Y1029 have been observed by others as autophosphorylation sites (Liu *et al.*, 1996; Salvatore *et al.*, 2000; Schuringa *et al.*, 2001). Y1090 and Y1096 are only present in the RET51 splice variant: the other 11 sites are present in all three splice variants. Phosphorylation of these sites leads to activation of mitogen-activated protein kinase (MAPK) pathways: rat sarcoma oncogene/extracellular signal-related kinase (RAS/ERK), p38 MAPK and ERK5; the janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway; the phosphatidylinositol-3-kinase (PI3K)/AKT pathway; a protein kinase C (PKC) pathway; and the c-Jun N-terminal kinase (JNK) pathway (Figure 1.4) (Ichihara *et al.*, 2004; Kodama *et al.*, 2005; Morandi *et al.*, 2011; Plaza-Menacho *et al.*, 2006; Santoro *et al.*, 2004). The pathways activated and the overall cellular response differs between the different RET ligands (Parkash and Goldman, 2009) and the different RET splice variants (Tsui-Pierchala *et al.*, 2002). Oncogenic mutations also alter the downstream signalling from RET (Liu *et al.*, 1996; Salvatore *et al.*, 2001).

## 1.4 RET biological function

Of the three RET splice variants RET9 and RET51 are the best characterised, are conserved amongst vertebrates, and have distinct but overlapping roles (Carter *et al.*, 2001). RET9 null mice show major developmental defects, primarily in the kidney and ENS, and generally die shortly after birth. RET51 null mice however are essentially normal (de Graaff *et al.*, 2001). Mice that express no functional RET isoform show even more severe developmental defects than the RET9 null mice, commonly showing kidney agenesis and a complete absence of enteric ganglia from the colon (Schuchardt *et al.*, 1994). RET, and particularly RET9, is therefore essential for proper development of the kidney and enteric nervous system in embryonic mice and this has also been shown to be the case in humans (Costantini and Shakya, 2006; Sasselli *et al.*, 2012).

More recently RET, along with GFRα3, has also been shown to be essential for the proper development of Peyer's patches, which are a major component of the lymphoid tissue associated with the gut (Veiga-Fernandes *et al.*, 2007). RET is also required for spermatogenesis. Impairment of RET signalling by mutation of phosphorylation site Y1062F (and L985P, a KD mutation of unknown impact) in RET9 in mice prevents spermatogonial differentiation and leads to apoptosis of spermatid precursor cells (Jain *et al.*, 2004).

As well as in the ENS and the excretory system, RET is also expressed in all other lineages of the peripheral nervous system and the motor and catecholaminergic neurons of the central nervous system (CNS) during early development. In adults RET expression is maintained in sympathetic and sensory neurons, the ENS, and the CNS motor, dopaminergic and noradrenergic neurons (Arighi *et al.*, 2005; Tsuzuki *et al.*, 1995). In adult dopaminergic neurons RET's ligand GDNF is an important survival factor, which is of particular interest since degeneration of these neurons is associated with Parkinson's disease. However GDNF treatment of Parkinson's disease clinically has so far produced inconclusive results (Hoffer and Harvey, 2011).

### 1.4.1 Conservation of RET function in evolution

A single RET gene is present in the mammalian, lower vertebrate and arthropod lineages of the animal kingdom (Shiu and Li, 2004). Uniquely the amphioxus genome

contains >100 RET pseudogenes, alongside a single canonical RET gene (D'Aniello *et al.*, 2008). RET is absent from all other animal kingdom phyla, seemingly lost in the course of evolution from the Echinoderms and Ascidians (Shiu and Li, 2004).

The model organisms zebrafish (*Danio rerio*), a lower vertebrate, and fruit fly (*Drosophila melanogaster*), an arthropod, have both been used to study RET. It has been shown in the zebrafish system that RET and GFL/GFRα function in the development of the ENS is conserved (Heanue and Pachnis, 2008; Marcos-Gutierrez *et al.*, 1997; Shepherd *et al.*, 2001; Shepherd *et al.*, 2004). It is therefore thought RET's biological function is conserved amongst vertebrates. In *Drosophila*, RET is expressed in an analogous range of tissue types as in the mammalian system (Hahn and Bishop, 2001) and therefore likely has an analogous function. Furthermore a *Drosophila* model of the RET-associated cancer syndrome multiple endocrine neoplasia type 2 (MEN2) has been developed, which, due to the conservation of signalling pathways (Abrescia *et al.*, 2005), has allowed kinome-wide drug profiling in the context of MEN2 oncogenic mutations, as well as direct studies of RET inhibitors (Dar *et al.*, 2012; Das and Cagan, 2010). However no GFL has been identified in *Drosophila* and therefore *Drosophila* RET remains an orphan receptor (Hatinen *et al.*, 2007).

## 1.5 RET and disease

RET has been associated with a number of diseases. Loss-of-function mutations in RET result in developmental disorders, such as Hirschsprung's disease and kidney agenesis, and gain-of-function mutations, aberrant expression or gene fusions/translocations can cause or promote tumorigenesis (Plaza-Menacho *et al.*, 2006).

### 1.5.1 Loss-of-function RET mutations and Hirschsprung's disease

Hirschsprung's disease is a developmental disorder of the ENS characterised by the enteric ganglia failing to colonise the entire length of the colon (Amiel *et al.*, 2008). This leads to a non-functional portion of the colon, which tonically contracts, resulting in intestinal obstruction. There are three different clinical subtypes of Hirschsprung's disease classified by the degree of aganglionosis: short-segment Hirschsprung's, where the rectum and a short portion of the colon are affected; long-segment Hirschsprung's,

where aganglionosis occurs in a longer stretch of the colon; and the rarer and more severe total colonic aganglionosis, where the entire colon is affected (Heanue and Pachnis, 2007). A number of genes have been linked to Hirschsprung's disease, with *RET* being the most common (Plaza-Menacho *et al.*, 2006; So *et al.*, 2011). Heterozygous mutations in *RET* have been observed in 50% of familial and 15–35% of sporadic Hirschsprung's patients (Heanue and Pachnis, 2007). A large number of Hirschsprung's-related loss-of-function RET mutations have been identified. These can be classified into five different classes based on the mechanism by which they confer loss-of-function (**Error! Reference source not found.**Plaza-Menacho *et al.*, 2006) (Pelet *et al.*, 1998; Plaza-Menacho *et al.*, 2006).

Class 1 mutations are found in the RET-ECD and result in decreased surface expression of RET by impairing proper folding of the protein and processing by the endoplasmic reticulum (ER) (Kjaer *et al.*, 2010; Kjaer and Ibanez, 2003b). Some of these affect calcium-binding or N-linked glycosylation sites (Kjaer and Ibanez, 2003b; Pelet *et al.*, 1998).

Class 2 mutations are a specialised group of class 1 mutations. They are all mutations of cysteine residues within the cysteine rich domain (CRD), leading to loss of an intramolecular disulphide bond, and leaving an unpaired cysteine. This results in increased misfolding, and decreased surface expression like the class 1 mutations. However these mutations also result in formation of disulphide-linked RET homodimers (Pelet *et al.*, 1998), which constitutively activates RET. These mutations are therefore both loss-of-function and gain-of-function mutations, and often result in a combined Hirschsprung's disease and MEN2A (multiple endocrine neoplasia type 2A, a cancer syndrome) syndrome, discussed in section 1.5.3 (Kjaer and Ibanez, 2003b).

Class 3 mutations are found in the KD, and confer loss-of-function by impairing catalytic activity (discussed in Knowles *et al.*, 2006). Class 4 mutations also affect the function, rather than cell surface expression of the receptor. These mutations affect the binding of downstream signalling molecules to phosphotyrosines in RET, leading to impairment of signalling downstream of RET (Geneste *et al.*, 1999).

### 1.5.2 Loss-of-function RET mutations and kidney agenesis

Commensurate with the well-studied role of RET in kidney development, RET mutations have been observed in some studies in up to 20–37% stillborn fetuses with uni- or bi-lateral kidney agenesis (Costantini and Shakya, 2006; Skinner *et al.*, 2008). However the importance/frequency of RET mutations in severe human kidney developmental defects has been questioned by a more recent study on a larger number of fetuses with bilateral defects. In this study RET mutations were only observed in 7% of cases, and largely did not appear to be causal, since they were shared with healthy parents (Jeanpierre *et al.*, 2011).

### 1.5.3 RET point mutations in the hereditary cancer syndrome MEN2

Germline gain-of-function point mutations in RET cause MEN2, of which there are three subtypes: MEN2A, MEN2B and familial medullary thyroid carcinoma (FMTC) (Arighi *et al.*, 2005; Donis-Keller *et al.*, 1993; Hofstra *et al.*, 1994; Mulligan *et al.*, 1993). MEN2 is characterised by medullary thyroid carcinoma (MTC), a malignant cancer, which develops from the calcitonin-secreting C-cells of the thyroid gland (Santoro *et al.*, 2004). MTC is the only pathology associated with FMTC, whereas MEN2A and MEN2B are associated with other cancers and pathologies, such as pheochromocytoma, a cancer of the adrenal gland. MEN2B is the most aggressive subtype, with widespread associated affects. These include ganglioneuromatosis of the gastrointestinal tract, medullated corneal nerves and marfanoid habitus (Ichihara *et al.*, 2004). The common RET mutations seen in MEN2 patients are detailed in Figure 1.6. Many of these have also been observed in sporadic MTC, with RET mutations found in around 50% of sporadic MTC cases (Arighi *et al.*, 2005).

The C634R mutation is the most common MEN2A mutation, observed in 50% of patients, with other C634 mutations accounting for a further 40% of MEN2A cases (Ichihara *et al.*, 2004). C634 mutations, as well as the other CRD cysteine mutations, results in the formation of constitutively active disulphide-linked homodimers of RET (Asai *et al.*, 1995; Santoro *et al.*, 1995). The proclivity of these cysteine mutants to produce RET homodimers has been shown to be linked to the self-association of the RET-TM. Dimerisation of RET via the TM brings the juxtamembrane regions of the

CRD, where all the MEN2A cysteine mutations are located, close together allowing intermolecular disulphide formation between the unpaired cysteines (Kjaer *et al.*, 2006).

Mutations in C634 and C630 do not affect RET folding and surface expression, and are therefore not linked to Hirschsprung's disease (Kjaer *et al.*, 2006). However mutations in C609, C611, C618 and C620 are class 2 Hirschsprung's mutations (see section 1.5.1), and can result in a joint MEN2A and Hirschsprung's disease syndrome (Plaza-Menacho *et al.*, 2006).

More rarely mutations within the RET-KD are observed in MEN2A, whereas the more severe MEN2B is primarily caused by KD mutations. The molecular basis for RET activation as a result of these KD oncogenic mutations is not well understood. M918T is associated with 95% of MEN2B cases, with a further 3–4% caused by A883F (Ichihara *et al.*, 2004; Smith *et al.*, 1997). The M918T mutation is highly activating, resulting in receptor autophosphorylation in the absence of ligand/co-receptor, and has been shown to be capable of mediating downstream signalling from the ER (Runeberg-Roos *et al.*, 2007). The mechanism underlying the activation of RET kinase activity by M918T is unclear. M918 resides in the substrate binding P + 1 loop of the RET-KD, and an analogous oncogenic mutation is also observed in MET RTK (M1250T). Based on the assumption that *in vivo* RET-KD adopts an inactive auto-inhibited conformation like that observed in non-phosphorylated crystal structures of its closest evolutionary relative, FGFR-KD, computational modelling showed the M918T mutation reduces the stability of this conformation (Dixit *et al.*, 2009). This inactive conformation of RET-KD has yet to be observed in a RET crystal structure however (Knowles *et al.*, 2006). The M918T mutation in the context of recombinant RET-ICD results in a 10× increase in ATP affinity, which is curious given the distance of M918 from the ATP binding pocket (Gujral *et al.*, 2006). M918T has also been proposed to confer activity by destabilising a *trans*-inhibitory interaction between RET-KDs (discussed further in section 1.7.2) (Knowles *et al.*, 2006). A further hypothesis for M918T's mechanism of action is that it promotes the detachment of the activation loop, which is required for *trans*-autophosphorylation of activation loop tyrosines by a partner KD. This activation loop detachment and engagement with a dimeric partner KD active site has been observed in crystal structures of a number of serine-threonine kinases (Oliver *et al.*, 2007)

### 1.5.4 RET fusion proteins in thyroid cancer, lung cancer and leukaemia

As well as point mutations, RET can also be activated by the production of fusion proteins, as observed in papillary thyroid cancer (PTC). PTC is the most common form of thyroid cancer (80–90% of thyroid cancers), and RET fusion proteins are associated with both sporadic and radiation induced PTCs (Santoro *et al.*, 2004). RET expression is not normally observed in normal thyroid follicular cells, and therefore these rearrangements result in the activation of RET expression as well as activity (Grieco *et al.*, 1990; Lam *et al.*, 1998). The first RET/PTC fusion protein was identified in 1990 (Grieco *et al.*, 1990), and there are now 11 known RET/PTC fusions (Figure 1.6) (Santoro *et al.*, 2006). These fusions link the RET-KD and C-tail to a constitutively dimeric fusion partner. This results in a cytoplasmic and constitutively active RET (Santoro *et al.*, 2006).

More recently similar RET fusions have been identified and proposed to be driver mutations in some lung adenocarcinomas (Kohno *et al.*, 2012; Lipson *et al.*, 2012; Takeuchi *et al.*, 2012) and in chronic myelomonocytic leukaemia (Ballerini *et al.*, 2012).

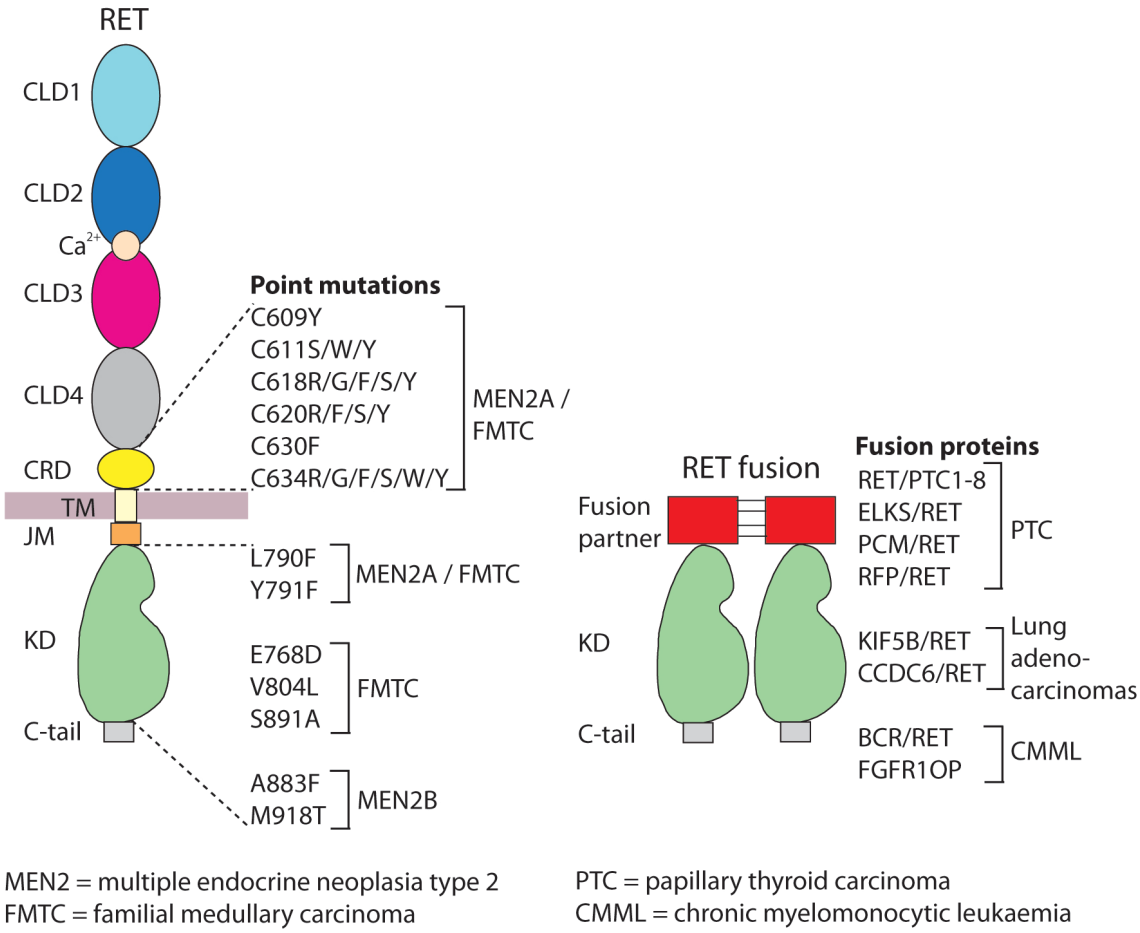



**Figure 1.6 Cancer-associated RET mutations and gene fusions**

Schematic showing some of the clinically observed oncogenic RET mutations based on the following citations: MEN2/FMTC (Kodama *et al.*, 2005; Plaza-Menacho *et al.*, 2006); PTC (Santoro *et al.*, 2006); Lung adenocarcinomas (Kohno *et al.*, 2012; Lipson *et al.*, 2012; Takeuchi *et al.*, 2012); CMML (Ballerini *et al.*, 2012).

### 1.5.5 Aberrant RET expression in breast, colorectal and other cancers

RET over-expression has been reported in many different cancers including breast, prostate, pancreatic, lung, bile duct, neural-crest derived cancers and melanoma (Morandi *et al.*, 2011). In many of these tissues RET is not normally expressed. More detailed analysis has been carried out on the role of RET in breast cancer, where over-expression has been observed specifically in oestrogen receptor (ER)-positive invasive breast cancers (Boulay *et al.*, 2008; Esseghir *et al.*, 2007; Plaza-Menacho *et al.*, 2010).

Recently RET has been proposed to play a tumour suppressor role in colorectal cancer. Aberrant *RET* gene methylation, which has been shown to result in decreased RET expression, has been observed in 27% of colon adenomas and 63% of colorectal cancers.

Furthermore restoration of RET expression in colorectal cell lines led to apoptosis (Luo *et al.*, 2012).

## 1.6 RET ectodomain structure

### 1.6.1 Cadherin-like domains predicted domain arrangement

The RET-ECD contains four cadherin-like domains (CLDs), originally identified by sequence alignment to members of the classical cadherin family, which revealed their distant similarity to extracellular cadherin (EC) domains (Figure 1.7) (Anders *et al.*, 2001; Schneider, 1992). RET is therefore an atypical cadherin superfamily member (Nollet *et al.*, 2000).

Cadherins are single-pass transmembrane proteins, involved in calcium-dependent cell adhesion. They are characterised by the presence of EC domain repeats in their ECDs. EC domains adopt a seven-stranded β-barrel fold (βA-βG), which RET CLDs share (Figure 1.7).

EC domains in vertebrate classical cadherins are arranged end to end, with their junctions rigidified by three bound calcium ions (Boggon *et al.*, 2002a; Harrison *et al.*, 2011; Shapiro *et al.*, 1995). The calcium-binding residues at EC–EC junctions are conserved. They are generally acidic residues and are found in five different motifs. Three motifs are from the preceding EC domain: E at the end of βA, DRE between βE and βF; and DXNDN in the EC–EC linker. And two motifs are from the succeeding EC domain: DXD after βB; and D at the end of βF (see Jin *et al.*, 2012 and Figure 1.7). Sequence analysis of the RET-ECD revealed that the CLD2–CLD3 junction contains almost the full complement of calcium-binding residues, which are conserved from human to *Drosophila* RET, lacking only the D at the end of βF in CLD3. RET is therefore predicted to bind calcium and display an end-to-end CLD2–CLD3 arrangement (Figure 1.7) (Anders *et al.*, 2001).

The RET CLD1–CLD2 and CLD3–CLD4 junctions contain only some of the conserved calcium-binding residues. CLD2 lacks both motifs (DXD and D) that would form the CLD1–CLD2 junction calcium-binding sites. CLD3 and CLD4 both contain some but not all of the calcium-binding residues that would form the CLD3–CLD4 junction

(Figure 1.7). The CLD1–CLD2 and CLD3–CLD4 junctions also contain longer linkers than the EC–EC junctions in classical cadherin and the CLD2–CLD3 junction. The lack of calcium-binding residues and the presence of longer linkers suggested that the arrangement of CLD1–CLD2 and CLD3–CLD4 is different to the calcium-bound EC domain arrangement of classical cadherin. This was found to be true for CLD1–CLD2, as discussed below (section 1.6.4) (Kjaer *et al.*, 2010). The arrangement of the CLD3–CLD4 junction was unknown, and is presented in this thesis.

(a)
A
B
C
D
E
F
G
Ca2+
α
β
C-cadEC1
C-cadEC2
C-cadEC3
C-cadEC4
C-cadEC5
CONSENSUS
hRET-CLD1
mRET-CLD1
rRET-CLD1
cRET-CLD1
xRET-CLD1
zRET-CLD1
hRET-CLD2
mRET-CLD2
rRET-CLD2
cRET-CLD2
xRET-CLD2
zRET-CLD2
hRET-CLD3
mRET-CLD3
rRET-CLD3
cRET-CLD3
xRET-CLD3
zRET-CLD3
hRET-CLD4
mRET-CLD4
rRET-CLD4
cRET-CLD4
xRET-CLD4
zRET-CLD4
Conserved residues: P Cadherin start consensus; E Calcium site; I β sheet hydrophobic; G Other conserved residue

**Figure 1.7 Alignment of RET CLDs to cadherin consensus sequences**

Alignment of vertebrate RET CLDs with human C-cadherin EC domains. The alignment is based predominantly on Anders *et al.*, 2001 and Jin *et al.*, 2012. Emphasis was placed on aligning cadherin consensus sequences, calcium-binding residues, and conserved hydrophobics, which define cadherin domain structure. C-cad = human (*Homo sapien*) C-cadherin; hRET = human RET; mRET = mouse (*Mus muscularis*) RET; rRET = rat (*Rattus norvegicus*) RET; cRET = chicken (*Gallus gallus*) RET; xRET = *Xenopus laevis* RET; zRET = zebrafish (*Danio rerio*) RET

### 1.6.2 The role of calcium in RET structure and function

RET has been shown to bind calcium directly (Anders *et al.*, 2001). And, as described above, the CLD2–CLD3 junction contains the cadherin consensus calcium-binding residues and is therefore predicted to be the calcium-binding site (Anders *et al.*, 2001). These calcium-binding residues are conserved from *Drosophila* to human RET.

Calcium is crucial for RET folding and intracellular processing as well as RET-mediated signalling. In the absence of extracellular calcium, or in the presence of a drug that causes the release of calcium from internal stores, RET fails to reach the cell surface but remains in the ER, presumably due to misfolding (van Weering *et al.*, 1998). Commensurate with this, mutations in the RET 'DRE', 'DEDD' and 'DXD' (R231H, D267N, D300N) calcium-binding motifs have been observed in Hirschsprung's disease, and lead to an impairment in RET intracellular trafficking (Pelet *et al.*, 1998).

Calcium is required for RET binding to GDNF, both *in vitro* and *in vivo* (Anders *et al.*, 2001; Nozaki *et al.*, 1998). In line with this RET autophosphorylation and activation of MAPK downstream signalling pathway in response to GDNF or Neurturin is calcium-dependent (Nozaki *et al.*, 1998).

These data suggest that the proposed calcium-conferred rigidity of the CLD2–CLD3 junction, analogous to the classical cadherin EC junctions, is vital for the proper folding and ligand-binding function of RET.

### 1.6.3 RET lacks an accessible adhesion motif present in cadherins

The presence of CLDs in the RET-ECD, which is unique for a RTK, led to speculation as to whether RET could mediate cell adhesion (Runeberg-Roos and Saarma, 2007). However this does not appear to be the case, except through binding GFL/GFRα's *in*

*trans*, since the elements involved in cadherins adhesive function are not conserved (Kjaer *et al.*, 2010; Nollet *et al.*, 2000).

Classical cadherins mediate cell adhesion by formation of a *trans* dimer between cadherin molecules from opposing cells. The N-terminal EC (EC1) domains from each monomer intimately interact, symmetrically swapping their N-terminal β-strands and burying the conserved tryptophan residue (W2) side chain in a conserved hydrophobic pocket in their partners EC1 domain (Figure 1.8a) (Harrison *et al.*, 2010). W2 is not conserved with RET, which instead has a phenylalanine. The non-classical T-cadherin also mediates cell adhesion by formation of a *trans* dimer, but this dimer is not formed by a strand-swap between EC1 domains. Rather the dimerisation interface spans both the EC1 and EC2 domains resulting in the formation of an X-shaped dimer, with the calcium-binding sites at the EC1–EC2 junctions juxtaposed at the centre of the X-shape (Figure 1.8b) (Ciatto *et al.*, 2010). The mechanisms by which other non-classical cadherins mediate cell adhesion have not yet been elucidated and may be different to those detailed above (Jin *et al.*, 2012).

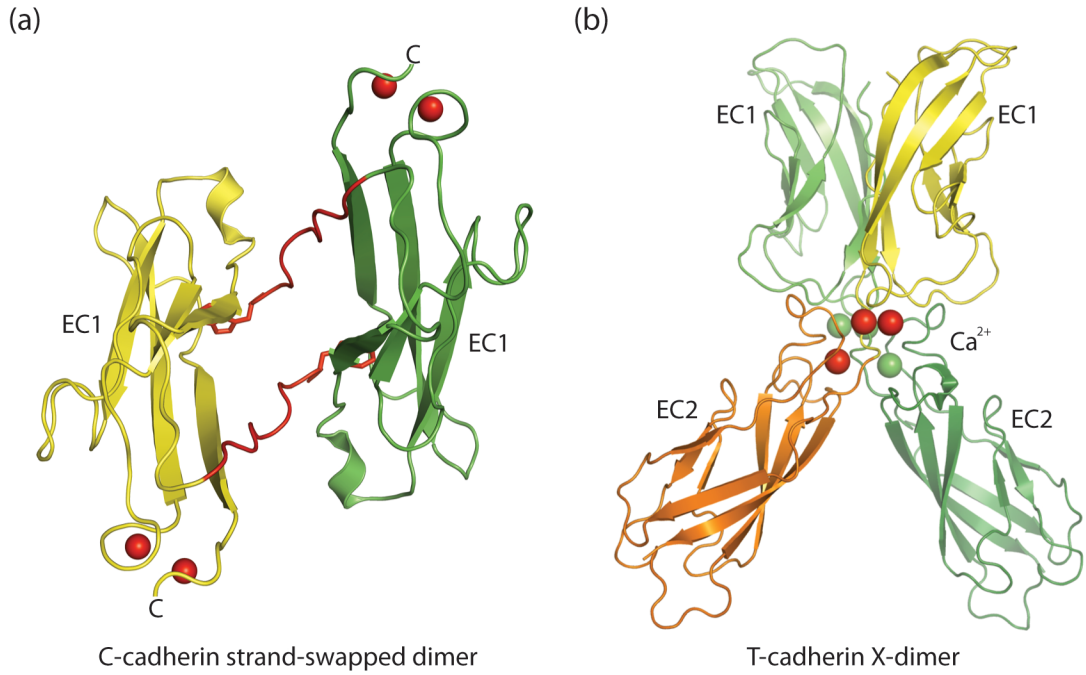



**Figure 1.8 Cadherins mediate cell adhesion via formation of *trans* dimers**

(a) Ribbon representation of the EC1 domains from a C-cadherin dimer (PDB 1L3W). EC1 domains from C-cadherin mediate symmetric *trans* dimerisation by a strand swap mechanism. The first β-strand (βA, coloured red) from each EC1 domain buries itself in its partner, an arrangement that is stabilised by the conserved W2 (side chain shown). (b) Ribbon representation of EC1–EC2 domains from a T-cadherin dimer (PDB 3K5S): one monomer is coloured yellow and orange with red calciums and the second is coloured green and dark green with green calciums. T-cadherin forms an X-shaped dimer, with the interface focussed around the EC1–EC2 calcium-binding junctions.

## 1.6.4 The crystal structure of human RET-CLD(1–2)

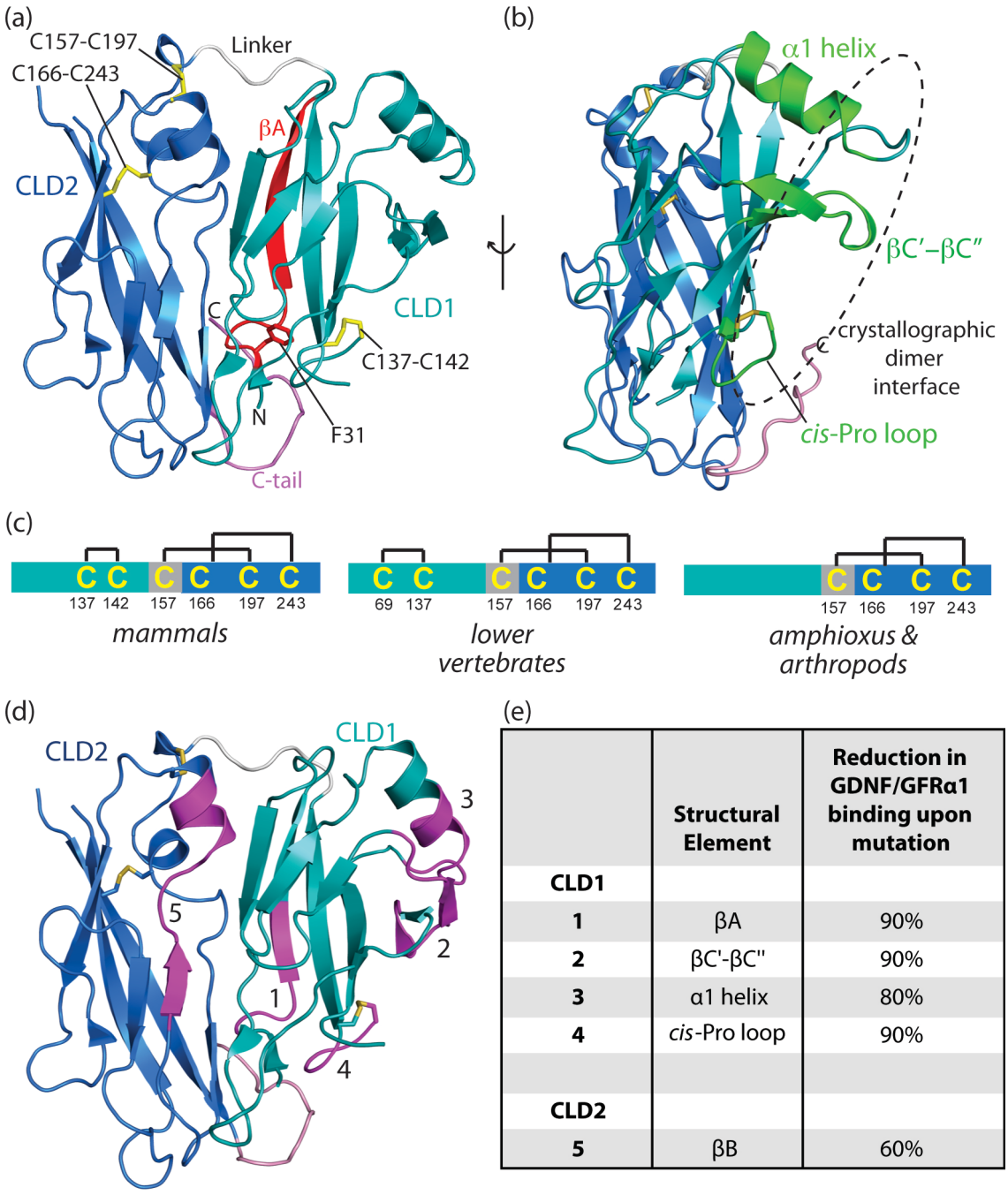
### *1.6.4.1 A clamshell CLD1–CLD2 arrangement*

The crystal structure of human RET-CLD(1–2) (hRET-CLD(1–2)) was solved by Dr. Svend Kjær in the Structural Biology Laboratory (Kjaer *et al.*, 2010). This revealed that CLD1–CLD2 adopts a clamshell arrangement similar to the calcium-free T-cadherin EC1–EC2 structure (Figure 1.9a) (Ciatto *et al.*, 2010). However unlike the calcium-free T-cadherin structure, this arrangement is not the result of the chelation of calcium from a normally linear calcium-rigidified junction, but rather is the constitutive hRET-CLD(1–2) arrangement. A disulphide bond formed between a cysteine in the CLD1–CLD2 linker and a cysteine in CLD2 (C157–C197) rigidifies the CLD1–CLD2 clamshell, and these cysteines are conserved from *Drosophila* to human. The interface

between CLD1 and CLD2 is predominantly hydrophobic and buries a surface area of ~1250 Å$^2$ (Kjaer *et al.*, 2010). The N-terminal β-strand of CLD1 is buried in the centre of this structure, packed into CLD1 (Figure 1.9a), thereby preventing the formation of strand-swapped *trans*-dimers as seen in classical cadherins (Figure 1.8a).

#### *1.6.4.2 Mammalian-specific CLD1 structural elements*

Both CLD1 and CLD2 have seven-stranded β-barrel folds, confirming their cadherin-like nature. The hRET-CLD(1–2) crystal structure also revealed that CLD1 contains three additional structural elements that are not shared with cadherins (Figure 1.9b). These are a short anti-parallel β-strand (βC'–βC"), an 11-residue α-helix (α1), and a six-residue loop containing a *cis*-proline and constrained by a disulphide bond (*cis*-Pro loop). These elements had already been implicated in RET/ligand/co-receptor complex formation (discussed further in section 1.6.5) (Kjaer and Ibanez, 2003a), and curiously are only conserved amongst mammalian RETs and not with lower vertebrate or arthropod RETs (Kjaer *et al.*, 2010). The βC'–βC" strand appears to be absent in non-mammalian RETs, which generally have a shorter loop between βC and βD. The α1 helix sequence is not conserved between mammalian and non-mammalian RETs. Finally the *cis*-Pro loop is absent from non-mammalian RETs, since one of the cysteines is not present in lower vertebrate RETs, and both are absent in arthropod RETs. In the lower vertebrate case an alternative disulphide appears to be present, formed between the remaining '*cis*-Pro' cysteine, and a lower vertebrate specific cysteine in the βC–βD loop, leading to an alternative structural arrangement at this CLD1 surface (Figure 1.9c) (Kjaer *et al.*, 2010).




| | Structural Element | Reduction in GDNF/GFRα1 binding upon mutation |
|---|---|---|
| **CLD1** | | |
| **1** | βA | 90% |
| **2** | βC'-βC'' | 90% |
| **3** | α1 helix | 80% |
| **4** | *cis*-Pro loop | 90% |
| | | |
| **CLD2** | | |
| **5** | βB | 60% |

**Figure 1.9 The human RET-CLD(1–2) crystal structure revealed mammal-specific structural elements and disulphide arrangements**

(a) The hRET-CLD(1–2) crystal structure (PDB 2X2U), shown in ribbon representation, reveals a clamshell arrangement with CLD1 (teal) folded back against CLD2 (blue). The N-terminal βA strand of CLD1 (red) is buried in CLD1. The disulphide bonds are shown as yellow sticks, and the C-tail, which in the absence of CLD3 is folded back onto CLD2, is coloured pink. (b) CLD1 contains additional mammal-specific structural elements not seen in classical cadherin structures, which are highlighted in green. (c) A schematic showing the different disulphide patterns observed in mammalian, lower vertebrate and arthropod/amphioxus RET-CLD(1–2). (d) hRET-

CLD(1–2) crystal structure with regions important for hGDNF/rGFRα1 binding, identified by homologue scanning mutagenesis, coloured magenta (Kjaer and Ibanez, 2003a). (e) The table details the reduction in hGDNF/rGFRα1 binding by hRET-ECD upon mutation of the identified regions to the corresponding *Xenopus* RET sequence.

#### *1.6.4.3 The hRET-CLD(1–2) dimer*

These mammalian-specific CLD1 structural elements, and particularly the loop between βC' and βC" and the *cis*-Pro loop, are principally involved in a dimeric interface observed in the hRET-CLD(1–2) crystal lattice (Figure 1.9d). These observations led to the proposal that the hRET-CLD(1–2) dimer may have some physiological relevance given its extent and the unusual interdigitation of mammal-specific arginine side chains at the interface. Whether the dimer is relevant to a ligand-free RET dimer or a ligand/co-receptor bound form remains to be shown. Data in Kjaer *et al.* suggests mutation of the interface interdigitating arginines can impede the formation of the RET ternary complex (Kjaer *et al.*, 2010).

### 1.6.5 Mapping ligand/co-receptor interactions onto RET

There have been two studies directly looking for the binding site for GFL/GFRα complexes on RET, which reached very different conclusions. The first, based on the observation that human RET cannot bind lower vertebrate (*Xenopus*) GDNF/GFRα1 and vice versa, used homologue scanning mutagenesis to identify the regions of human RET that were required for human/*Xenopus* RET chimeras to recognise hGDNF/rGFRα1 (Kjaer and Ibanez, 2003a). They concluded that human RET-CLD1, CLD2 and CLD3 were required. In further analysis they showed that certain predicted surface exposed regions of CLD1 and CLD2 were important for the GDNF/GFRα1 interaction, since mutation of these stretches in human RET to the equivalent *Xenopus* sequence significantly reduced binding to hGDNF/rGFRα1 (Kjaer and Ibanez, 2003a). The regions they identified are shown on the subsequently solved hRET-CLD(1–2) crystal structure in Figure 1.9d&e (Kjaer *et al.*, 2010). Three of these regions are found at the hRET-CLD(1–2) crystallographic dimer interface, which is consistent with the conclusion of (Kjaer *et al.*, 2010); that this CLD(1–2):CLD(1–2) interaction occurs in the context of GFL/GFRα binding and stabilises the complex (see section 1.6.4.3).

The second study looking at the RET GFL/GFRα binding interface, was a cross-linking-based study (Amoresano *et al.*, 2005). This group identified three interactions, based on cross-linked peptides, between RET, GFRα1 and GDNF that were not observed in the GDNF/GFRα1 complex alone. These were all focused on the same region of GFRα1-D1, which cross-linked to RET-CRD, the base of RET-CLD4 and the N-terminal tail of GDNF. They did not identify an interaction between RET and GFRα1-D2 or D3, which is the primary RET binding surface (Scott and Ibanez, 2001). The caveat to these results is that the cross-linked residues they identified between GDNF and GFRα1 in the absence of RET, are on the same side of the GDNF/GFRα1 crystal structure, but are not directly interacting. In this study they also show in cell-based assays that RET-CLD(1–3) cannot compete with RET for GDNF/GFRα1 binding, showing that RET-CLD4CRD is important in RET's interaction with GDNF/GFRα1 (Amoresano *et al.*, 2005).

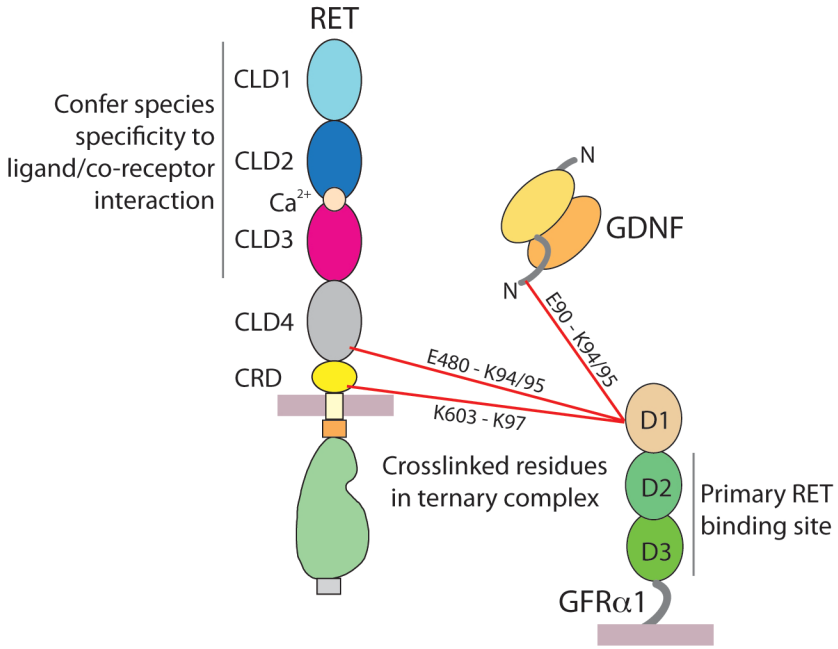



**Figure 1.10 Formation of the RET/ligand/co-receptor ternary complex**

Schematic showing what is currently known about the interactions between RET and ligand/co-receptor (Amoresano *et al.*, 2005; Kjaer and Ibanez, 2003a; Parkash *et al.*, 2008). For simplicity the primary interaction between GDNF and GFRα1-D2, which occurs in the presence and absence of RET, is not shown.

## 1.7 Regulation of RET's tyrosine kinase activity

Binding of GFL/GFRαs to the RET-ECD results in the activation of RET's intracellular tyrosine KD and autophosphorylation, which leads to the activation of downstream signalling. The molecular mechanisms underlying the regulation of RET's kinase activity are not fully determined.

### 1.7.1 RET regulation by activation loop autophosphorylation

RET has two autophosphorylation sites in its activation loop—Y900 and Y905. In the context of MEN2A mutation C634F, which results in constitutively active RET dimers, mutating Y905 to a phenylalanine abolishes cell transforming activity and inhibits, though does not abolish, both RET autophosphorylation and phosphorylation of an exogenous substrate *in vitro* (Kato *et al.*, 2002). Y900F in the MEN2A context has no effect on transforming ability in cells or on *in vitro* kinase activity. However the Y900F Y905F double mutant shows even less activity than Y905F alone *in vitro* (Kawamoto *et al.*, 2004). These data suggest a role for activation loop phosphorylation in the regulation of RET kinase activity and activation of downstream signalling.

However recombinant RET-KD and RET-ICD show only a 3–4× increase in *in vitro* kinase activity towards an exogenous peptide substrate following autophosphorylation, which is modest in comparison to the 10–200× increases observed for known autophosphorylation-regulated RTKs (Knowles *et al.*, 2006). In this context the Y905F mutant had the same catalytic activity as non-phosphorylated wild type RET-KD, linking the modest increase in kinase activity following autophosphorylation to Y905 phosphorylation (Knowles *et al.*, 2006). The crystal structure of non-phosphorylated RET-KD revealed an active, catalytically competent, conformation, with the same global conformation as the Y905 phosphorylated RET-KD structure (Figure 1.11a&b) (Knowles *et al.*, 2006). Therefore the non-phosphorylated RET-KD structure did not reveal an *in cis* mechanism of inhibiting kinase activity, and implied that activation loop phosphorylation could not fully account for regulation of RET kinase activity.

Understanding the RET autoinhibitory mechanism(s) is crucial for understanding how oncogenic mutations subvert RET autoinhibition. The lack of an identified *in cis* mechanism within the KD for inhibiting RET kinase activity led us to look both at the

other elements within the ICD and whether they play a role in kinase inhibition, and also at a potential *in trans* inhibitory mechanism, discussed below.

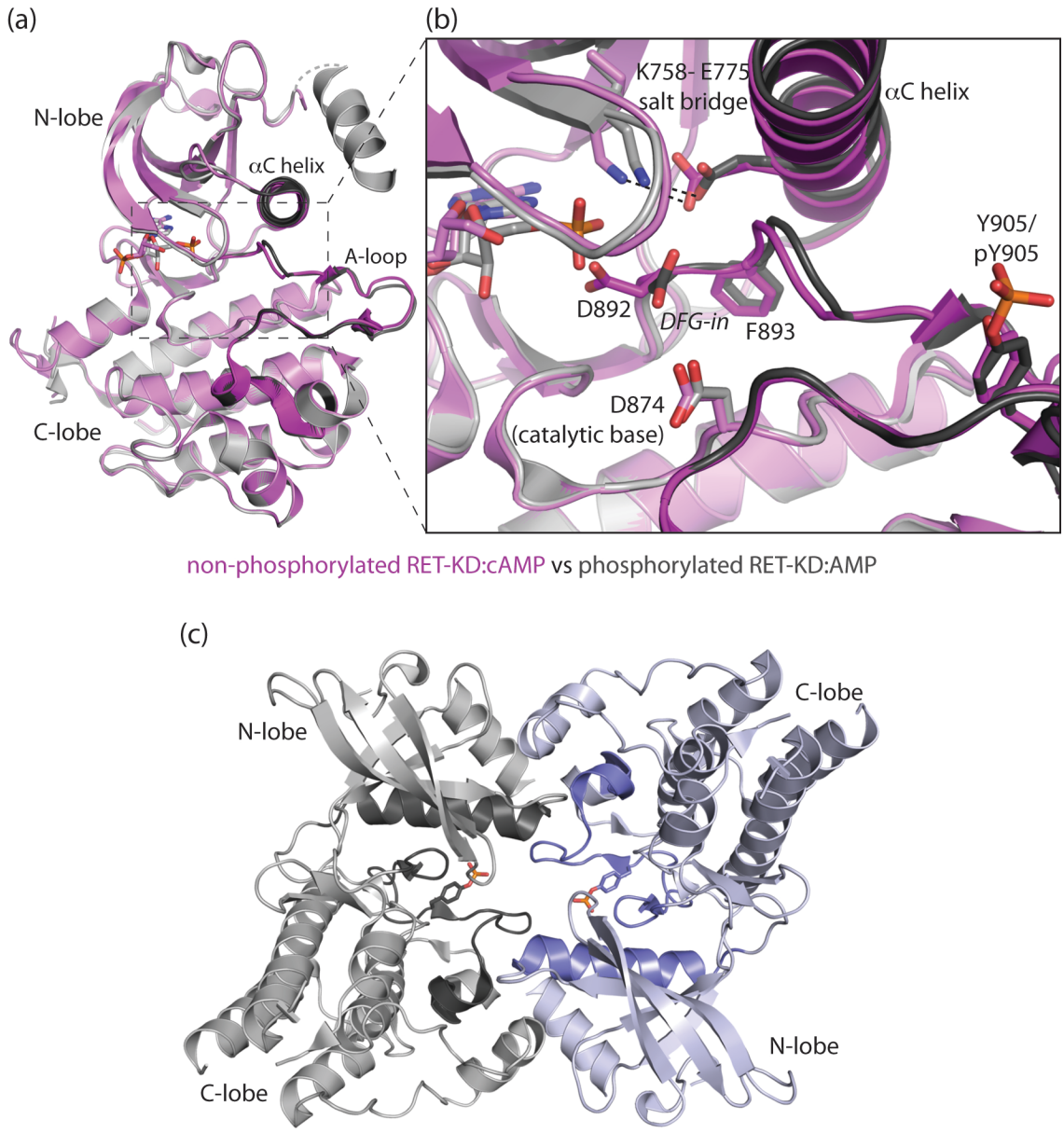



**Figure 1.11 Structural analysis of RET-KD activation mechanisms**

(a) Superposition of the non-phosphorylated RET-KD: cyclic adenosine monophosphate (cAMP) crystal structure (pink ribbon, with magenta activation loop and αC helix; PDB 2IVS) and the pRET-KD:AMP structure (grey, with dark grey activation loop and αC helix; PDB 2IVT). The bound nucleotides are shown in stick representation. (b) Close up of the active site, with side chains of key catalytic and regulatory residues shown in stick representation. (c) Head-to-tail dimer of pRET-KD:AMP, as seen in the crystal, showing the occlusion of the substrate-binding site. Second protomer is coloured pale blue and with slate activation loop and αC helix.

### 1.7.2 *Trans*-inhibition of RET-KD: a putative RET regulatory mechanism

Some proportion of RET has also been proposed to dimerise in the absence of ligand, since its TM is able to self-associate. This self-association, though not required for normal RET activation by GFL/GFRα's, is the basis for the formation of disulphide linked RET dimers in MEN2A-associated cysteine mutants (also discussed in section 1.5.3) (Kjaer *et al.*, 2006). If ligand-independent dimerisation occurs, inhibitory mechanisms must then be present to prevent misfiring of the kinase in the dimer in the absence of ligand.

All published RET-KD crystal structures contain a symmetric head-to-tail dimer within the crystal lattice (Figure 1.11c) (Knowles *et al.*, 2006; Mologni *et al.*, 2010). The asymmetric unit contains either a single molecule, in which case the dimer is crystallographic (e.g. PDB 2IVT), or the asymmetric unit contains a dimer and the two protomers are related by non-crystallographic two-fold symmetry (e.g. PDB 2IVS). Formation of this dimer buries a surface area of 1090–1380 $Å^2$ (10% of the total surface area) involving 43 residues from each protomer. Given the proclivity of RET-KD to form this dimer in the high concentration environment of the crystal and the significant size of the dimer interface it was hypothesised that it could be a relevant *in vivo* (Knowles *et al.*, 2006). This dimer buries the cleft between the N and C lobes containing the active site; and the substrate-binding P+1 loop, at the C-terminus of the activation loop, is intimately involved in the dimer interface.

Formation of this dimer *in vivo* would occlude the substrate-binding site and prevent kinase activity—a *trans*-inhibitory mechanism (Knowles *et al.*, 2006). It was therefore hypothesised that formation of this dimer could be part of the mechanism for regulating kinase activity. In line with this some MEN2 oncogenic mutation sites, including the highly transforming MEN2B M918T, are located at the dimer interface, and could therefore increase RET activity by destabilising this dimer (Knowles *et al.*, 2006). A similar dimer is observed in the apo FGFR2-KD crystal structure (PDB 2PSQ), but not in the substrate peptide bound structure (PDB 2PVF) (Chen *et al.*, 2007). This is particularly curious since phylogenetically the FGFR family is RET's closest RTK relative (Manning *et al.*, 2002b).

An autoinhibitory KD dimer, observed in crystal structures of inactive EGFR-KD, has similarly been proposed to contribute to the inhibition of kinase activity in ligand-free EGFR dimers (Endres *et al.*, 2011; Jura *et al.*, 2009). This dimer is stabilised by the EGFR C-tail, and in support of this dimer being relevant physiologically, loss of the C-tail is oncogenic (Pines *et al.*, 2010).

### 1.7.3 The potential role of the RET JM and C-tail in kinase regulation

The role of the ICD elements flanking the RET-KD in kinase regulation is not well characterised. The RET-JM contains one autophosphorylation site (Y687) and the C-tail contains three that are present in all RET isoforms (Y1015, Y1029, Y1062) (Kawamoto *et al.*, 2004; Liu *et al.*, 1996; Salvatore *et al.*, 2000; Schuringa *et al.*, 2001). However none of these phosphorylation sites have as yet been shown to involved in kinase regulation. The JM is implicated in RET kinase regulation, and contains a protein kinase A phosphorylation site (S696) which modulates RET activity (Asai *et al.*, 2006). A small number of MEN2 and sporadic MTC associated mutations have also been identified in the JM, suggesting it normally plays an inhibitory role (Cordella *et al.*, 2006; Runeberg-Roos and Saarma, 2007).

## 1.8 RET kinase inhibitors and cancer

### 1.8.1 Targeting tyrosine kinases in cancer

Deregulation of a tyrosine kinase is a common feature of many different cancers and tyrosine kinases have been a popular target for development of cancer therapies. The success of imatinib, a small molecule ATP pocket binding inhibitor of ABL kinase, in treating chronic myeloid leukaemia (Druker *et al.*, 2001), showed that successful therapies could be discovered through rational design rather than simple serendipity (Cohen, 2002). As well as ATP pocket binding small molecule inhibitors, other strategies for targeting tyrosine kinases have also been tried, including protein substrate mimics, remote allosteric pocket binding inhibitors and RTK-ECD binding antibodies, minibodies and apatamers (Cowan-Jacob *et al.*, 2009). Although thus far only ATP pocket binding inhibitors and antibodies have been successfully applied in the clinic

(Cowan-Jacob *et al.*, 2009), but an allosteric inhibitor of ABL is also showing some promise (Zhang *et al.*, 2010).

### 1.8.2 RET is a validated target for cancer therapy

MTC is recalcitrant to standard radiotherapy and chemotherapy. Surgery to remove the thyroid can be curative if the cancer has not metastasised. There was no approved treatment for non-resectable advanced or metastatic MTC until 2011 when vandetanib (ZD6474), a RET and VEGFR2/3 inhibitor, was approved by the FDA with a 44% objective response rate in the phase III trial (Thornton *et al.*, 2012). Vandetanib was originally discovered as potent RET oncogenic mutant inhibitor (Carlomagno *et al.*, 2002), and targets the ATP pocket of the RET-KD (Knowles *et al.*, 2006).

Due to the success of vandetanib, targeting RET in MTC treatment is validated. However vandetanib is not curative, prolonging advanced MTC survival by just 6 months over placebo (22 months over 16 months). It is also associated with a number of adverse reactions. Treatment was discontinued in 12.1% of vandetanib-treated patients versus 3.0% of placebo-patients due to adverse reactions and 2% of vandetanib-treated patients died as a result of adverse reactions (Thornton *et al.*, 2012).

### 1.8.3 RET inhibitors

#### *1.8.3.1 ATP pocket binding inhibitors*

There are a number of known small molecule RET inhibitors, which fall into two categories. Type 1 inhibitors bind in the ATP pocket, in particular they occupy the hydrophobic cleft where the adenine ring of ATP binds and make hydrogen bonds with residues in the linker between the N and C-lobes (the kinase hinge), which line the back of the ATP pocket (Liu and Gray, 2006). These inhibitors are generally compatible with the active conformation of the kinase, with the kinase conserved DFG motif at the start of the activation loop, adjacent to the nucleotide-binding site, in the catalytically competent *DFG-in* conformation (Liu and Gray, 2006). The crystal structures of the RET-KD bound to a number of type 1 inhibitors such as PP1 and vandetanib have been solved, showing that these inhibitors are compatible with the active RET *DFG-in* conformation (Knowles *et al.*, 2006; Mologni *et al.*, 2010).

Type 2 inhibitors specifically target the inactive *DFG-out* conformation, binding both in the ATP pocket and the neighbouring hydrophobic pocket, which is occupied primarily by the phenylalanine of the DFG motif when in the *DFG-in* conformation (Liu and Gray, 2006). Sorafenib is an example of a RET type 2 inhibitor (Carlomagno *et al.*, 2006), however no crystal structures of RET bound to a type 2 inhibitor have been published.

Due to the high similarity between the structures of active kinases, achieving kinase specificity with type 1 inhibitors is difficult, leading to many off-target effects. It was thought that targeting the inactive forms of kinases would allow more specific inhibitors to be developed, although the type 2 inhibitors target an inactive conformation (*DFG-out*) that is shared amongst many kinases. One problem with targeting the inactive conformation of kinases is that it opens the possibility for inhibitor resistance mutations that further activate the kinase, by inhibiting the inactive conformation (Johnson, 2009).

#### *1.8.3.2 Sunitinib and RET*

Sunitinib (Sutent®) is a type 1 inhibitor. It inhibits RTKs PDGFR, KIT, FLT3, VEGFR and RET and is a potent anti-tumour and anti-angiogenic agent (Chow and Eckhardt, 2007; Mendel *et al.*, 2003). It has been approved for the treatment of renal cell carcinoma and imatinib-resistant gastrointestinal stromal tumours. *In vitro* sunitinib inhibits VEGFR2 with a $K_i$ of 9 nM and PDGFR-β with a $K_i$ of 8 nM. Similarly the IC50 for sunitinib's inhibition of RTK autophosphorylation following ligand stimulation in cells is 10 nM for both VEGFR2 and PDGFR-β (Chow and Eckhardt, 2007).

Sunitinib has been shown to inhibit oncogenic RET autophosphorylation in cultured human MTC cells with an IC50 of ~100 nM (Chow and Eckhardt, 2007). Because of this sunitinib has been trialled in a small number of patients with advanced MTC or PTC, including a phase II clinical trial for MTC, with some success (35% response rate in the phase II trial) (Bajetta *et al.*, 2009; Bugalho *et al.*, 2009; Carr *et al.*, 2010; Higgins *et al.*, 2009; Kelleher and McDermott, 2008; Lanzi *et al.*, 2009).

Despite being a type 1 inhibitor, binding predominantly in the ATP pocket, the crystal structure of KIT-KD bound to sunitinib surprisingly showed KIT-KD in an inactive

conformation with *DFG-out*. This crystallographic finding was corroborated by *in vitro* kinase assays showing that sunitinib preferentially inhibits the inactive form of KIT-KD (Gajiwala *et al.*, 2009). This is a similar situation to imatinib and ABL kinase, where the crystal structure of the ABL:imatinib complex also surprisingly revealed imatinib binding to an inactive *DFG-out* conformation of ABL (Nagar *et al.*, 2002).

#### *1.8.3.3 Other types of RET inhibitor*

The only RET inhibitors that have been used clinically are ATP pocket binding inhibitors. A RET-ECD targeted monoclonal antibody and a dimerisation inhibiting aptamer (D4) have however been developed and used in cells (Plaza-Menacho *et al.*, 2006).

## 1.9 Project aims

The overall aim of my thesis was to investigate and define the molecular architecture, protein:protein interactions and regulation of the RET receptor tyrosine kinase. The conformational changes underlie RET's activation in response to ligand/co-receptor are unknown, as was the architecture of the entire RET-ECD. In this thesis I sought to address these questions as follows:

To elucidate the architecture of the RET-ECD, I employed small-angle X-ray scattering (SAXS) to produce a model for the entire RET-ECD architecture, as described in Chapter 2. This chapter also describes progress towards determining the crystal structure of CLD(1–4).

To investigate the basis for ligand/co-receptor recognition by the RET-ECD, I carried out the first single particle analysis electron microscopy (EM) 3D reconstruction for a mammalian RET ternary complex, as described in Chapter 3. From this I describe an interpretation of the reconstruction for the arrangement of RET, ligand and co-receptor in the RET ternary complex, based on known structures of the components, which is consistent with some of the published functional data.

Biochemical and structural data have revealed mammal-specific structural elements important for human RET's engagement with ligand/co-receptor. In order to explore how a lower vertebrate RET is able to recognise an essentially identical ligand/co-

receptor, I reconstituted and purified a lower vertebrate RET ternary complex for future structural analysis, as described in Chapter 4. Preliminary negative stain EM data suggests the lower vertebrate RET ternary complex has a similar overall architecture to the mammalian complex.

Current structural understanding of the regulatory mechanisms governing RET's tyrosine kinase activity is limited since the RET-KD has only been observed crystallographically in one conformation in a single dominant crystal lattice. To develop our understanding, I sought to obtain the crystal structure of the RET-KD in a different crystal lattice arrangement. In Chapter 5 I present a high-resolution structure of the RET-KD obtained during these efforts, which revealed some previously unseen features and flexibility of the RET-KD. Alongside this I also sought to determine the mode-of-action of sunitinib, a clinically relevant chemical inhibitor of RET, by solving the structure of sunitinib bound to the RET-KD, which is also presented in Chapter 5.

# Chapter 2. Structural analysis and architecture of the RET ectodomain

## 2.1 Chapter aims

The aim of this chapter was to produce a model for the RET-ECD architecture using experimental data from SAXS and X-ray crystallography. The RET-ECD is comprised of four CLDs, followed by a CRD. Previous studies in the laboratory defined the clamshell arrangement for the CLD(1–2) region of human RET and gave evidence of disulphide swapping within this region between lower and higher vertebrate RET (Kjaer *et al.*, 2010). However, there is little data available concerning the relative position and interdomain contacts between the CLD2 and CLD3, between CLD3 and CLD4, as well as between CLD4 and the CRD regions of RET. A model for the RET-ECD was a pre-requisite step before embarking on an EM analysis of the RET ternary complex with GDNF and GFRα co-receptor, since it would allow us to interpret the arrangement of the components in the EM 3D reconstruction map.

Although efforts to crystallise the entire RET-ECD were unsuccessful, towards the end of the thesis crystals were obtained for CLD(1–4) from a lower vertebrate (zebrafish) RET. An X-ray diffraction dataset was collected and underwent a preliminary analysis. In parallel, high quality SAXS data were measured at Synchrotron SOLEIL, Paris. These data provided a description of the arrangement of the CLDs within the RET-ECD with an unexpected kink between CLD3 and CLD4. SAXS analysis of CRD deletion constructs revealed the location of CRD, packed against the CLD4. Furthermore production of CLD4CRD constructs highlighted a possible role for the CLD(1–3) in the proper arrangement of the CRD, consistent with the SAXS data and previous unpublished observations by Dr. Svend Kjær. Finally circular dichroism confirmed the CRD contains limited regular secondary structure, with its apparently globular fold as observed by SAXS defined by its 7 intramolecular disulphides.

## 2.2 RET-ECD sequence alignment and structural conservation

I undertook structural characterisation of the RET-ECD using a lower vertebrate (zebrafish) RET-ECD in parallel to Dr. Svend Kjær's work on the human RET-ECD. Amino acid sequence alignment of RET-ECDs from three mammalian and three lower vertebrate RETs (Figure 2.2) shows fairly high sequence divergence, with 41% sequence identity between human and zebrafish RET-ECD (excluding the signal sequence; Table 2.1). The highest sequence similarity is seen in the membrane proximal CRD, which shows 50% identity between human and zebrafish with complete conservation of 14 of the 16 cysteine residues in the human CRD. The greatest divergence is between the membrane distal CLD(1–2), with just 35% identity between human and zebrafish. Despite this, human and zebrafish RET-ECD are thought to adopt the same overall fold.

| | Human vs. Rat | Human vs. Zebrafish |
|---|---|---|
| **CLD1** | 80% | 36% |
| **CLD2** | 76% | 34% |
| **CLD3** | 80% | 38% |
| **CLD4** | 74% | 48% |
| **CRD** | 84% | 50% |
| **ICD** | 93% | 84% |

**Table 2.1 Sequence identity between individual RET subdomains from vertebrate RETs**
Amino acid sequence identities between individual RET subdomains from human (*Homo sapien*) RET and another mammalian RET (rat: *Rattus norvegius*) and a lower vertebrate RET (zebrafish: *Danio rerio*).

As shown in Figure 1.7, CLD(1–4) from both human and zebrafish align well to cadherin consensus elements including the 7 β-strand pattern, and are therefore predicted to share a 7-strand β-barrel fold. Furthermore, as discussed in section 1.6.4.2, the clamshell arrangement of CLD(1–2) is likely conserved between human and zebrafish due to the conservation of the disulphide tether between the CLD1 linker and CLD2. The calcium binding residues at the CLD2–CLD3 junction are also conserved, as is the long linker between CLD3 and CLD4, making it likely that the overall arrangement of the CLDs is conserved. The CRD contains a conserved pattern of 14 cysteines between human and zebrafish, which likely define the fold of this domain.

The similarity between mammalian and lower vertebrate CLD4 and the CRD is further demonstrated by the ability of human-*Xenopus* RET chimeras containing *Xenopus* CLD4CRD to bind human ligand/co-receptor, despite CLD4CRD being involved in and required for ligand/co-receptor binding (Amoresano *et al.*, 2005; Kjaer and Ibanez, 2003b).

Differences between human and zebrafish RET-ECD structure are likely different embellishments on the same overall fold. Mammalian-specific structural elements have been identified in CLD1, based on the human CLD(1–2) crystal structure, which are important for ligand/co-receptor binding (Kjaer *et al.*, 2010). One further notable difference between zebrafish and all the other vertebrate RETs shown in Figure 2.2, is an 8 amino acid insertion in CLD4 between predicted strands βC and βD, which is also present in RET from other fish.

Structural characterisation of the zebrafish RET-ECD would therefore provide a description of the overall RET-ECD architecture, most likely shared with mammalian RET, as well as revealing the structure of the lower vertebrate specific features providing insights into the differences in ligand/co-receptor recognition between mammalian and lower vertebrate RET. A number of different constructs including the whole zebrafish RET-ECD and a number of smaller fragments were produced in the course of this thesis for structural studies (Figure 2.1). The production and characterisation of these are discussed in this chapter.

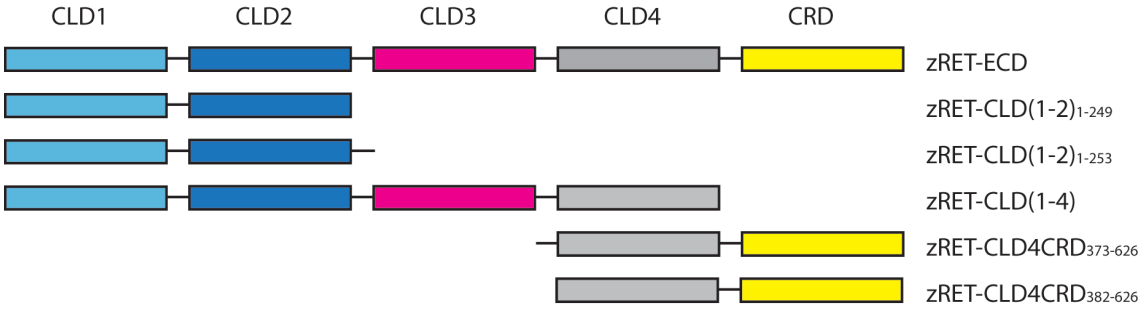



**Figure 2.1 zRET-ECD constructs examined in this thesis**

Schematic of the zebrafish RET-ECD constructs produced as part of this thesis. Individual predicted domains are depicted as boxes with the linkers shown as lines.

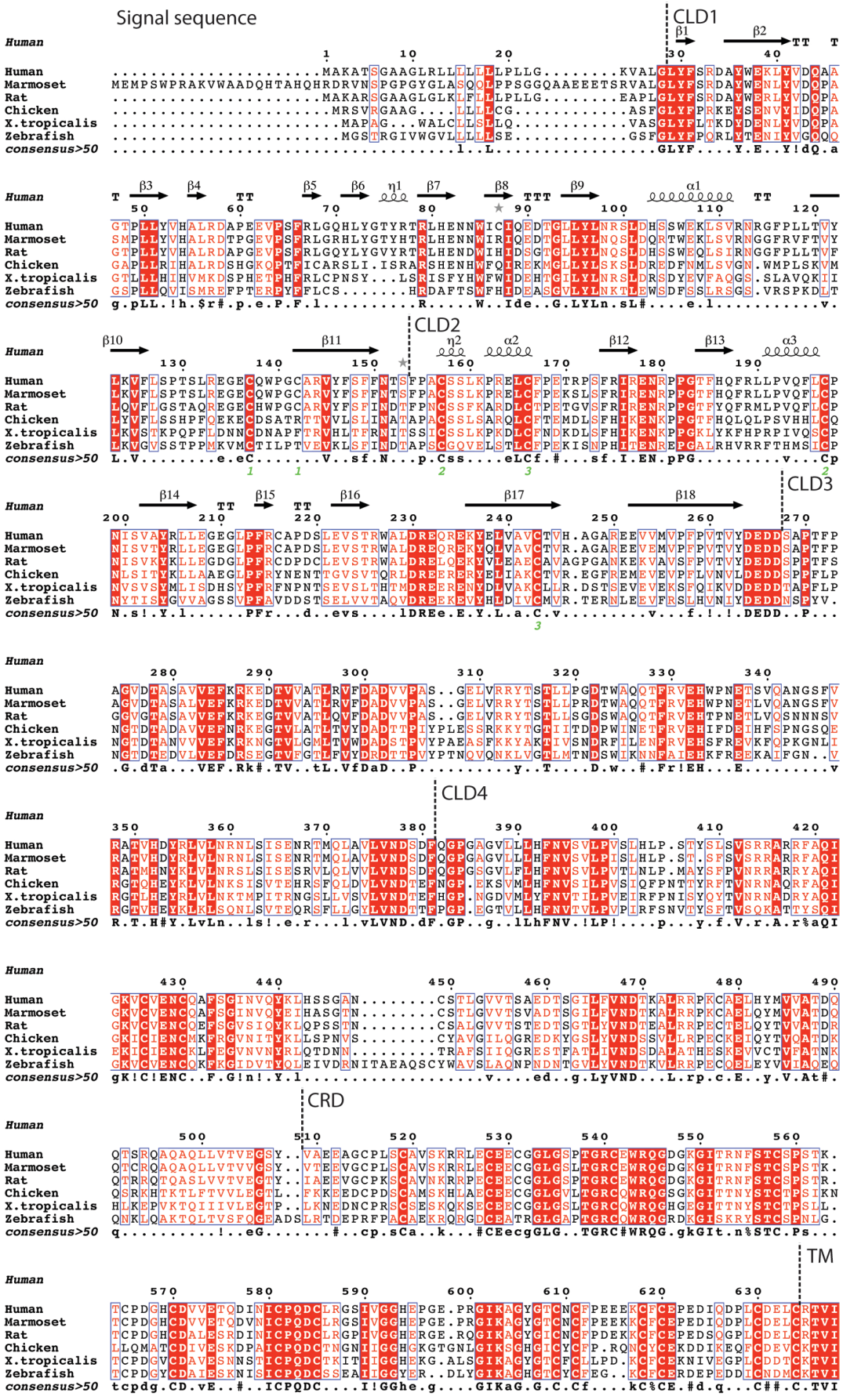

Signal sequence
CLD1
CLD2
CLD3
CRD
TM
Human
Marmoset
Rat
Chicken
X.tropicalis
Zebrafish
consensus>50

**Figure 2.2 Alignment of vertebrate RET-ECD amino acid sequences**

A CLUSTALW2 alignment of human (*homo sapien*), marmoset (*Callithrix jacchus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*), *Xenopus tropicalis* and zebrafish (*Danio rerio*) RET-ECD protein sequences. Secondary structure elements observed in the human RET-CLD(1–2) crystal structure (PDB 2X2U) are annotated above the alignment. Disulphide-linked pairs of cysteine residues seen in the structure are marked with green numbers beneath the alignment. Completely conserved residues are boxed and highlighted in red with white text, and similar residues boxed with red text. The consensus amino acid sequence is displayed beneath the alignment, absolutely conserved residues are capitalised. Domain boundaries indicated above the alignment are based on the alignment of the CLDs to cadherin EC domains (Anders *et al.*, 2001).

## 2.3 RET-ECD protein production

### 2.3.1 Rationale for using the baculovirus system for protein expression

To produce recombinant forms of RET-ECD for structural analysis, I prepared zebrafish RET-ECD (zRET-ECD) using a recombinant baculovirus to infect an insect cell host. Prior to this thesis, considerable efforts were made to produce recombinant human RET-ECD (hRET-ECD) in bacteria, insect cells and mammalian cells. Only the mammalian cell expression was successful (Kjaer and Ibanez, 2003b). After the crystal structure of hRET-CLD(1–2) was determined, it was rationalised that the presence of a tightly constrained *cis*-Pro loop within CLD1 and two unpaired cysteines may have contributed to the difficulty in producing recombinant hRET-ECD (Kjaer *et al.*, 2010). We hypothesised that producing a lower vertebrate RET lacking the unpaired cysteines and without the *cis*-Pro loop found in higher vertebrates may allow protein production in insect cells. This eukaryotic expression host offers an ER and Golgi apparatus for the folding, glycosylation and export of membrane and extracellular proteins, which bacterial systems do not possess, whilst adding less complex glycans to extracellular proteins than unmodified mammalian cells meaning the resulting glycoproteins may crystallise more readily (Jarvis *et al.*, 2009).

### 2.3.2 RET-ECD expression cassette development

For zRET-ECD production, I engineered a construct containing the coding sequence for the entire zRET-ECD, ending at the final cysteine (C626) of the CRD (the zRET clone was kindly provided by Dr. Tiffany Heanue and Prof. Vassilis Pachnis). This was cloned into a baculovirus transfer vector within an expression cassette based on that used in the production of the hRET-CLD(1–2); utilising the native secretion signal sequence, and encoding a C-terminal Protein A tag (Kjaer *et al.*, 2010).

The linker between the RET-ECD and Protein A coding sequences initially just contained the TEV protease recognition site, however TEV cleavage of the resulting fusion protein was not efficient. I therefore extended the linker by addition of the 3C-protease recognition site and three glycine residues (RET-*3C*-*TEV*-Protein A). RET-ECD could then be cleaved efficiently from its tag by both TEV and 3C-proteases. All the proteins produced using this construct contain a short C-tail that is a cloning/protease site artefact following cleavage from the Protein A tag. In the 3C-protease cleaved proteins used exclusively in this thesis, this tail is 10 amino acids long (GSEFLEVLFQ).

Baculoviruses generated from this transfer vector were used for protein expression in *Spodoptera frugiperda* 9 (Sf9; Invitrogen) and *Trichopulsia ni* derived High Five™ (Hi5; Invitrogen) cells cultured in suspension. Initially cells were cultured in grace's media, supplemented with fetal bovine serum. However yields were much improved, and more reliable, after transferring to serum-free media. The best protein expression yields were obtained in Hi5 cells grown in serum-free SfIII media (Invitrogen), with the media harvested 112 hours after infection.

### 2.3.3 Large scale zRET-ECD production in insect cells

Large-scale expression of zRET-ECD in Hi5 cells, infected with baculovirus at a multiplicity of infection (MOI) of 2, produced up to 4 mg protein per litre of cells. Affinity purification of the Protein A tagged material from the harvested media was carried out using IgG-sepharose resin (1ml/l culture). The zRET-ECD was eluted by 3C-protease cleavage of the Protein A tag, and then subject to further purification by size exclusion chromatography and occasionally anion exchange chromatography.

The zRET-ECD protein was stable throughout purification. It eluted as a single peak from a size-exclusion column, and the eluted protein was highly pure by Coomassie-stained SDS-PAGE (Figure 2.3a). The size-exclusion purified material bound to an anion exchange HiTrap Q column and could be eluted with increased salt concentration. About half the material eluted in a sharp peak with the remainder coming off as a broad shoulder at higher salt (Figure 2.3b&c). The basis for this elution profile is unknown, but it implies that the protein is heterogeneous with regard to its charge.

The size exclusion purified protein was soluble to > 25 mg/ml in a simple buffer without glycerol or detergent, and was stable at 4 °C for at least two weeks. The protein was therefore suitable for structural studies by X-ray crystallography and SAXS.

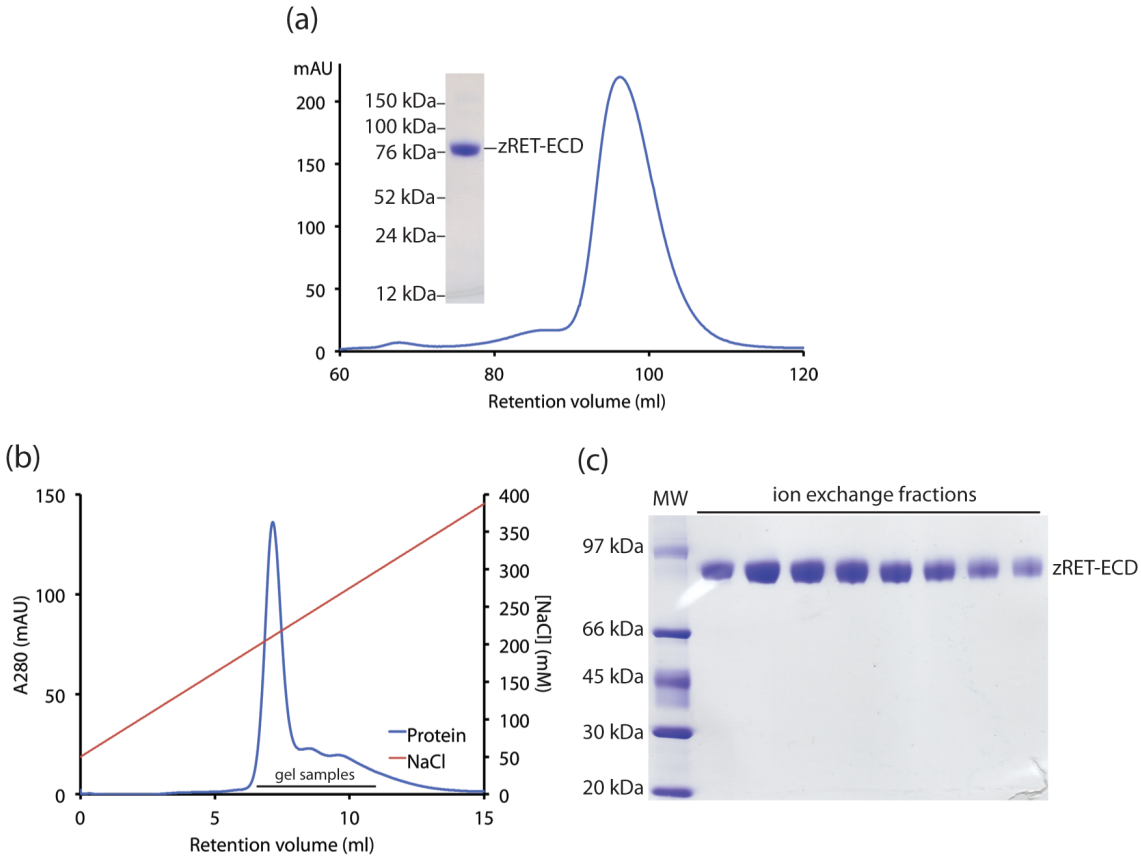



**Figure 2.3 zRET-ECD purification**

(a) Size exclusion chromatography of affinity purified zRET-ECD using a Superdex 200 column. Inset: Non-reducing SDS-PAGE of size exclusion peak material. (b) Anion exchange chromography of size exclusion purified zRET-ECD on a HiTrap Q column. (c) Non-reducing SDS-PAGE of anion exchange fractions from across the ion exchange chromatogram. MW = molecular weight markers.

### 2.3.4 Unsuccessful efforts to produce zRET-CLD(1–2)

As a direct comparison to the hRET-CLD(1–2), I engineered a zRET-CLD(1–2) construct. It was hoped that this protein would be reasonably simple to prepare and to crystallise, since we predicted it would display the same overall compact globular shape as the human protein. My interest in CLD(1–2) was stimulated by the hope that a crystal structure of zRET-CLD(1–2) would allow us to determine how the structure of a lower vertebrate RET is different to the human RET. In particular I was keen to see the structural impact of the swapped disulphide and altered loop regions that prevent lower vertebrate RET interacting with human ligand/co-receptor complexes (Kjaer and Ibanez, 2003a). A zRET-CLD(1–2) crystal structure would also reveal whether a similar crystallographic dimer to that observed for hRET-CLD(1–2) is present in the crystal lattice in the absence of the mammalian-specific elements, which mediate dimerisation in the human structure (Kjaer *et al.*, 2010).

Two different zRET-CLD(1–2) constructs were produced: one ending exactly where the final beta strand of hRET-CLD2 terminates (zRET-CLD(1–2)$_{1–249}$); and the second including the four acidic calcium-binding residues from the CLD2–3 linker (zRET-CLD(1–2)$_{1–253}$; Figure 2.1). Baculoviruses expressing these constructs utilising the RET-*3C*-*TEV*-Protein A expression cassette, produced usable levels of soluble protein in insect cells (0.5 mg/l). However SDS-PAGE and size exclusion chromatography of the affinity-purified protein showed that the protein produced from both constructs tended to self-aggregate. The protein showed many higher molecular weight contaminating bands on SDS-PAGE (Figure 2.4a) and, consistent with this, size exclusion chromatography showed a broad elution profile consistent with overlapping monomer, dimer and higher molecular weight peaks (Figure 2.4b). This was the case for both constructs.

The poor quality of the produced zRET-CLD(1–2)$_{1–249}$ was also apparent in a thermal melt assay. No proper melt curve could be observed for this protein in a range of buffers, consistent with the protein not being properly folded at the starting temperature of 20 °C. This unexpected result led to a decision not to pursue this construct further.

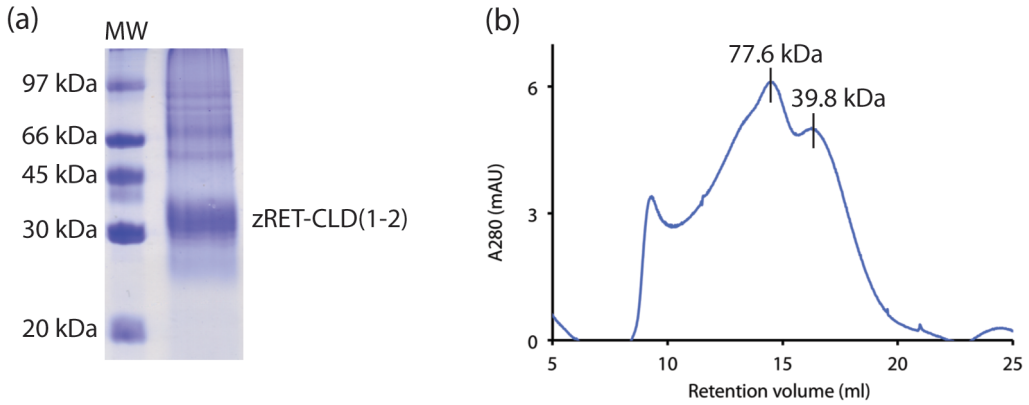



**Figure 2.4 zRET-CLD(1–2) purification**

(a) Non-reducing SDS-PAGE of affinity purified zRET-CLD(1–2)$_{1–249}$. (b) Size exclusion chromatography of affinity purified material using a Superdex 200 column.

### 2.3.5 Successful zRET-CLD(1–4) protein production

A construct containing just the four CLDs of zebrafish RET was also produced. This construct was originally generated in order to aid interpretation of zRET-ECD envelope produced by SAXS analysis (detailed in section 2.5.2). It was also hoped to be a good crystallographic target since it is made up of structured domains connected by fairly short linkers.

zRET-CLD(1–4) was expressed and purified in the same way as the complete zRET-ECD. The protein yield from Hi5 cells was about the same as the zRET-ECD—up to 4 mg/l in serum-free media.

Like zRET-ECD, zRET-CLD(1–4) was well behaved and stable. It eluted as a single peak from size exclusion chromatography, and the protein from the peak was pure by SDS PAGE and highly soluble (up to ~40 mg/ml). zRET-CLD(1–4) eluted with a slightly different profile to zRET-ECD from the same anion exchange column. As for zRET-ECD the trace shows a dominant sharp peak, but the shoulder eluting at higher salt is less broad and appears to consist of a single peak (Figure 2.5). Again the basis for this is unclear since the protein corresponding to both peaks is indistinguishable by SDS-PAGE.

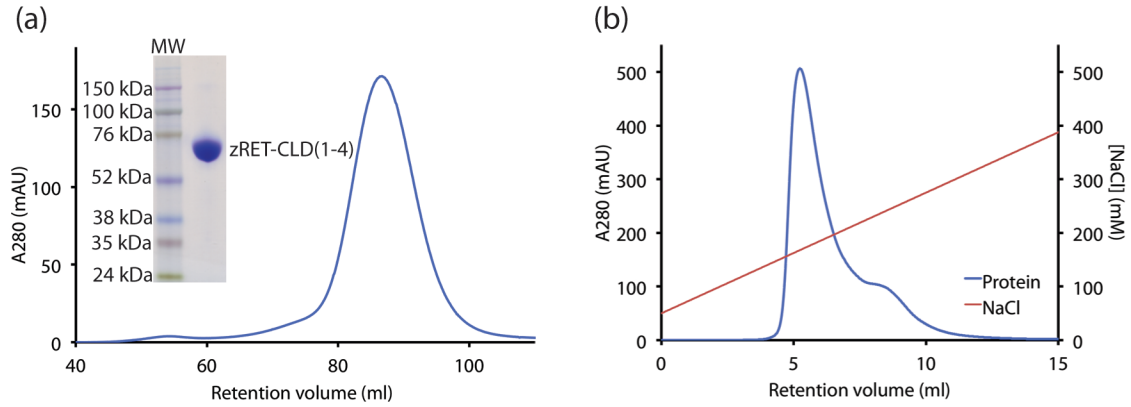



**Figure 2.5 zRET-CLD(1–4) purification**

(a) Size exclusion chromatography of affinity purified zRET-CLD(1–4) using a Superdex 200 column. Inset: Non-reducing SDS-PAGE of size exclusion peak material. (b) Anion exchange chromatography of size exclusion purified zRET-CLD(1–4) on a HiTrap Q column.

## 2.4 Recombinant RET-ECD biophysical characterisation

### 2.4.1 zRET-ECD is monomeric in solution

zRET-ECD elutes as a single peak from size exclusion as if it were 130–150 kDa, based on column calibration with molecular weight standards. Dynamic light scattering (DLS) of the size exclusion purified material, confirms it is monodisperse and gives an estimated molecular weight of 125–140 kDa. Similarly by native PAGE the dominant species runs at 150 kDa, with a small proportion running as a dimer of this species at around 300 kDa (Figure 2.6). Although since native PAGE separates proteins on the basis of both size and intrinsic charge the molecular weight values from reference to the molecular weight standards may not be a true reflection of the molecular weight of the protein. By SDS-PAGE the protein runs as a monomer at around 85 kDa—much closer to its predicted molecular weight of 80.8 kDa (69.3 kDa protein plus 11.5 kDa sugar). These results imply that the zRET-ECD could be a dimer in solution.

However size exclusion coupled multi-angle light scattering (SEC-MALS), gave an estimated molecular weight of 83.5 kDa (Figure 2.11 and Table 2.5). Since this technique uses the differential refractive index to measure concentration based on standard values for proteins, the molecular weight of glycosylated proteins will be over-estimated due to the different refractive index of the glycans. Nonetheless the estimated

molecular weight is consistent with the monomeric molecular weight by SDS-PAGE, and not with a dimeric molecular weight of 140–160 kDa. The protein is therefore predominantly monomeric in solution, with a small dimeric portion as shown by native PAGE. The overestimated molecular weight by size exclusion and DLS are likely due to the elongated shape of zRET-ECD, expected based on other tandem cadherin domain proteins.

The molecular weight estimate across the size exclusion peak by SEC-MALS is reasonably constant, showing the material is homogeneous at the resolution of this method. The low level of dimer present in the purified material as seen by native PAGE shows that the protein does self-associate although only with low affinity. This is likely why the SAXS data, as discussed later, is improved by running the protein on a size exclusion column directly before data collection, since this should separate out the dimeric species (see section 2.7).

All the biophysical properties described above for zRET-ECD also hold for zRET-CLD(1–4). As for the zRET-ECD it appeared to be much larger than its predicted molecular weight of 57 kDa plus glycosylation by size exclusion, native PAGE and dynamic light scattering. However its estimated molecular weight by MALS was 67.5 kDa, again in much better agreement with a monomeric species.

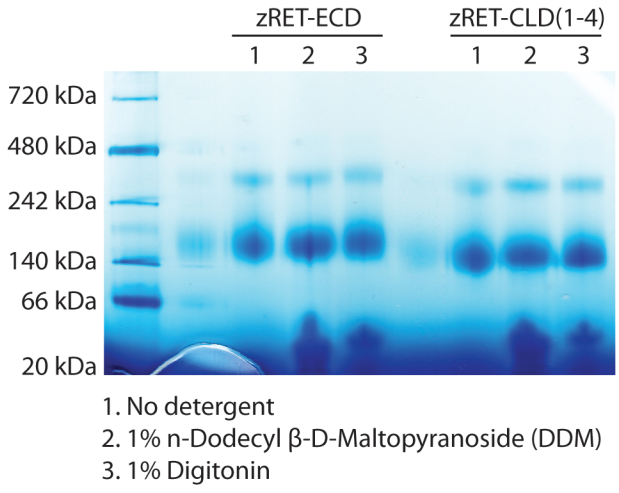



**Figure 2.6 Native PAGE of zRET-ECD proteins shows low-level dimerisation**

Coomassie stained native PAGE of size exclusion purified zRET-ECD and zRET-CLD(1–4). Addition of 1% DDM or digitonin detergents does not alter the level of the dimeric species.

### 2.4.2 Recombinant zRET-ECD is properly folded and functional

Circular dichroism of both the recombinant zRET-ECD and zRET-CLD(1–4) confirmed the proteins were folded and chiefly β-sheet in character, as expected for proteins consisting of predominantly cadherin-like domains (data shown in section 2.10.4). Furthermore zRET-ECD binds immobilised zebrafish ligand/co-receptor in the presence of calcium (data shown in section 4.3.2). These results confirm that the recombinantly produced zRET-ECD proteins are properly folded and able to perform their native function.

## 2.5 Preliminary structural studies of the RET-ECD

### 2.5.1 Crystallisation screening of zRET-ECD

Since the zRET-ECD could be produced at reasonable levels from insect cells, and the protein following purification was pure, highly soluble, homogeneous and properly folded, it was therefore suitable for crystallisation screening. The size exclusion purified zRET-ECD was subject to extensive crystallisation screening at both 4 and 20 °C, using a range of protein concentrations (3.5–20 mg/ml), salt concentrations (100–500 mM NaCl) and with or without 5% glycerol in the protein buffer. The ion exchange step was also added to the purification scheme. However no protein crystals were produced. It was therefore clear that crystallisation of this protein would be non-trivial, and further work pursuing crystallisation would require changes to the construct and/or glycan content. A successful strategy is discussed later in this chapter.

### 2.5.2 Preliminary efforts to produce a SAXS model of zRET-ECD

In order to obtain some understanding of the architecture of the RET-ECD whilst crystal trials were ongoing, the low-resolution technique, SAXS was used. This technique provides information about the intramolecular distances within a macromolecule in solution and can therefore be used to generate a model of the overall shape of a macromolecule, without the need for crystals. The first round of SAXS experiments, which informed later experiments with improved sample handling and data collection procedures, are discussed in this section.

#### 2.5.2.1 SAXS data collection at the European Synchrotron Radiation Facility (ESRF)

SAXS data were collected over a range of protein concentrations from 2–7 mg/ml for both the zRET-ECD (produced by myself) and the hRET-ECD (produced in mammalian cells by Dr. Svend Kjær) at the ID14-3 beamline, ESRF. Samples were flowed continuously through a quartz capillary mounted in vacuum for data collection, such that fresh sample was exposed to the X-ray beam for each frame of data collected. Raw X-ray scattering images from the detector for each frame of data were then converted at the beamline to 1D SAXS profiles of intensity of the scattered beam (I) as a function of the modulus of the scattering vector (s). Ten frames of data were collected for each protein concentration and were compared. The frames with superimposable SAXS data profiles, and therefore not showing any radiation damage effects, were combined and averaged to improve the signal-to-noise ratio. Later frames of data from the highest protein concentration (~7 mg/ml) of both samples showed radiation-induced aggregation. This was indicated by the increased scattered beam intensity at low angles compared with the earlier frames, which corresponds to an increase in the longest distances within the sample in the beam. These frames were therefore discarded before averaging of the data from this protein concentration.

Comparison between the averaged SAXS data curves from the four different concentrations of both proteins showed some evidence of association effects (or aggregation) at the higher protein concentrations (> 5.5 mg/ml). This was based on the radius of gyration (Rg) of the sample in the beam at each concentration, determined using the Guiner approximation (Guinier and Fournet, 1955) in PRIMUS (Konarev *et al.*, 2003), which showed a slight upward trend at increased concentrations.

#### 2.5.2.2 ESRF SAXS data processing

The SAXS data from the lower (2–5 mg/ml) protein concentrations of both zRET-ECD and hRET-ECD were then processed using the ATSAS (Konarev *et al.*, 2006) suite of SAXS data processing programs. GNOM (Svergun, 1992) was used to fit the SAXS data and generate a pair distance distribution function of vectors within the sample in

the beam, based on an iteratively determined maximum distance of 200 Å (Table 2.2; Figure 2.7).

| Protein | Rg (Å) | Rg error (Å) | Dmax (Å) |
|---|---|---|---|
| **zRET-ECD** | 51.3 | 0.13 | 200 |
| **hRET-ECD** | 49.0 | 0.31 | 200 |

**Table 2.2 Zebrafish and human RET-ECD SAXS data processing statistics**

Processing of SAXS data with GNOM. The table details the maximum vector length (Dmax) used in fitting the data and the fit-derived real space radius of gyration (Rg).

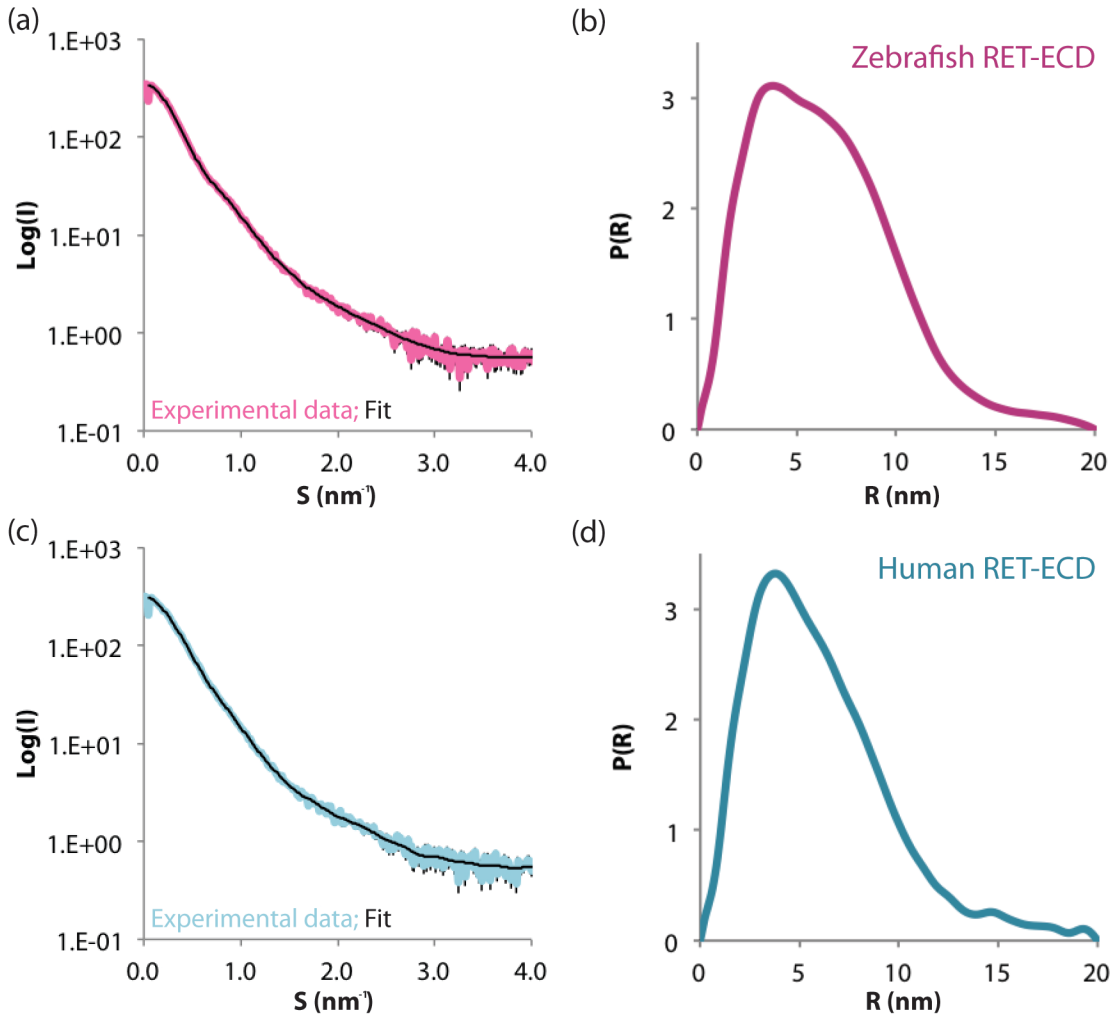



**Figure 2.7 Zebrafish and human RET-ECD SAXS data processing**

Processing the SAXS data collected at ESRF with GNOM to produce pair distance distribution plots. (a) zRET-ECD (2 mg/ml) experimental SAXS data curve (pink, with black error bars) compared with GNOM processing fit (black line). s = modulus of the scattering vector, I = intensity of the scattered beam at S. (b) Distance distribution plot derived from fit to the SAXS data. R = vector distances within the sample, P(R) = pair distance distribution function. (c) & (d) hRET-ECD (2 mg/ml).

#### *2.5.2.3 zRET-ECD and hRET-ECD ab initio models*

The GNOM processed data were then used to generate 20–30 *ab initio* models of zRET-ECD and hRET-ECD using DAMMIF (Franke and Svergun, 2009). These models were then overlaid and the consistent features used in the generation of a consensus *ab initio* model using DAMAVER (Volkov and Svergun, 2003) (Table 2.3).

| Protein | NSD mean | NSD deviation |
|---|---|---|
| **zRET-ECD** | 0.871 | 0.039 |
| **hRET-ECD** | 0.844 | 0.033 |

**Table 2.3 SAXS *ab initio* model generation statistics for human and zebrafish RET-ECD**

Generation of *ab initio* molecular envelopes with DAMMIF. The table details the mean correspondence (normalised spatial discrepancy, NSD) between the 20–30 generated models for each protein. Using DAMAVER these models were then overlaid and the common features were used in generation of the final models.

The consensus *ab initio* models of zRET-ECD and hRET-ECD, showed pleasing similarity, and exhibit a curved but extended conformation (Figure 2.8). This suggests the overall architecture of the lower and higher vertebrate RET-ECD is the same in solution. The only difference between the *ab initio* models is the presence of a longer tail on the zebrafish model, creating an overall 'question-mark' shape.

Fitting the CLD(1–2) crystal structure and 2 cadherin domains, representing CLD3 and 4, into the *ab initio* envelopes was ambiguous, since the envelopes were apparently too large for the predicted ECD subdomain structure and the CRD is the same size as a CLD. One possible interpretation with CLD(1–2) at the top of the models (as depicted) with CLD3 and CLD4 arranged end to end is illustrated in Figure 2.8 b&d. However there are other interpretations consistent with these envelopes. CLD4 could be accommodated folded back on CLD3, like the CLD(1–2) arrangement. Furthermore the components would also fit in the alternative orientation with the CLD(1–2) structure at the bottom. Therefore the *ab initio* models, while similar, were not conclusive, and did not allow a reliable assignment of RET-ECD subdomains.

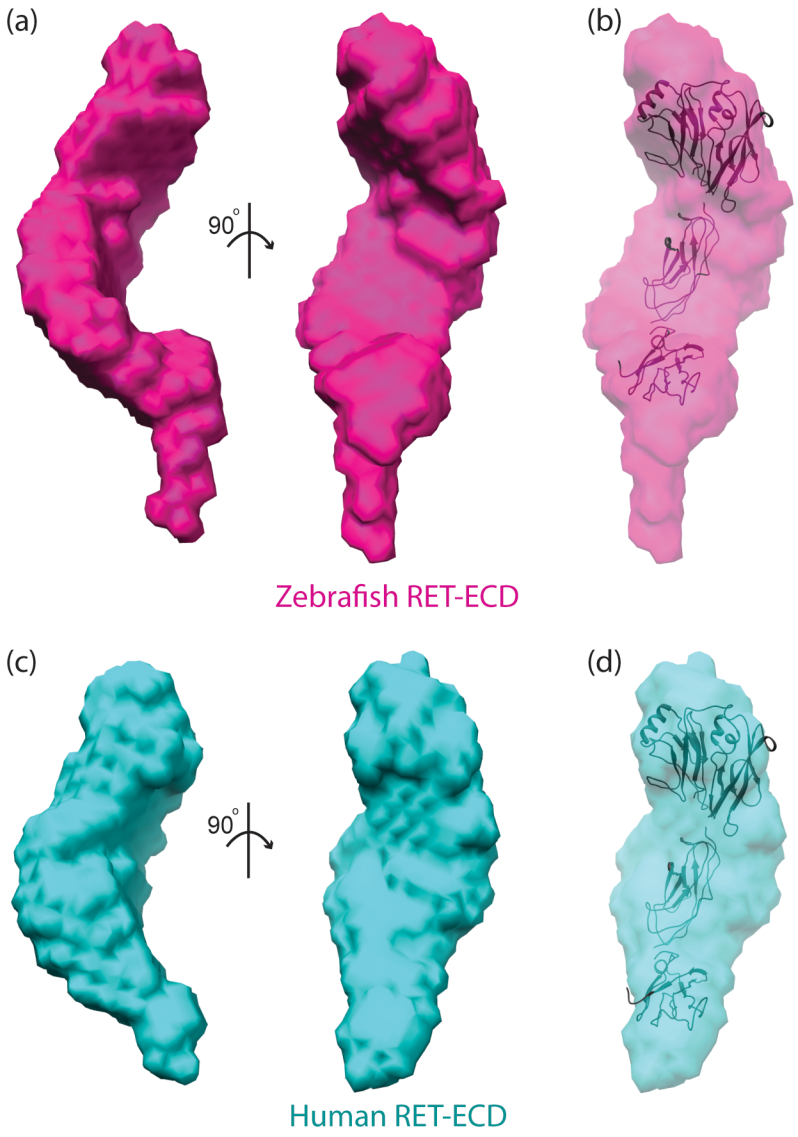



**Figure 2.8 *Ab initio* SAXS models of zebrafish and human RET-ECD**

Consensus SAXS *ab initio* molecular envelopes of (a) zRET-ECD and (c) hRET-ECD. One possible arrangement of the CLD domains (ribbon representation) is shown fitted into the (b) zebrafish and (d) human *ab initio* molecular envelopes.

#### *2.5.2.4 Approaches taken to improve the SAXS data*

In order to improve the SAXS data such that a reliable interpretation of resulting *ab initio* envelopes could be made, three avenues were taken. Firstly I began to pursue protein deglycosylation to reduce the potential impact of the sugar groups on the *ab initio* model (section 2.6). Secondly, as the higher concentrations of protein showed the

material had the propensity to aggregate consistent with the native PAGE results, and since even very low levels of aggregation can seriously impact the SAXS data, we applied for time at the SAXS beamline (SWING) at Synchrotron SOLEIL. This beamline has an online high pressure liquid chromatography (HPLC) system, allowing data collection on material directly after elution from a size exclusion column. Thus largely eliminating aggregation from the protein sample entering the data collection capillary (section 2.7). Finally, to define the orientation of the SAXS models—i.e. the N and C termini—I produced the zRET-CLD(1–4) construct, which lacks the C-terminal CRD (section 2.3.5).

## 2.6 Deglycosylation

Given the failure of crystallisation screening of the zRET-ECD and zRET-CLD(1–4), as well as the overlarge molecular envelope of zRET-ECD from SAXS analysis, deglycosylation of zRET-ECD and zRET-CLD(1–4) was pursued. Glycosylated proteins can be difficult to crystallise since the glycans attached to glycoproteins are flexible, heterogeneous and often prevent the close packing of glycoproteins required for crystal formation (Chang *et al.*, 2007). I therefore hypothesised that this could be the reason why the zRET-ECD proteins were not crystallising. The impact of glycosylation on SAXS data is not well studied, but their removal would clarify whether the size of the *ab initio* envelopes is purely the impact of aggregation in the sample or partly the influence of the glycans decorating the protein. If the glycans were found to have an impact, deglycosylation would then provide more interpretable SAXS data. For these reasons I sought to reduce the sugar content of the zRET-ECD proteins, firstly by enzymatic deglycosylation and subsequently by site-directed mutagenesis of predicted N-linked glycosylation sites.

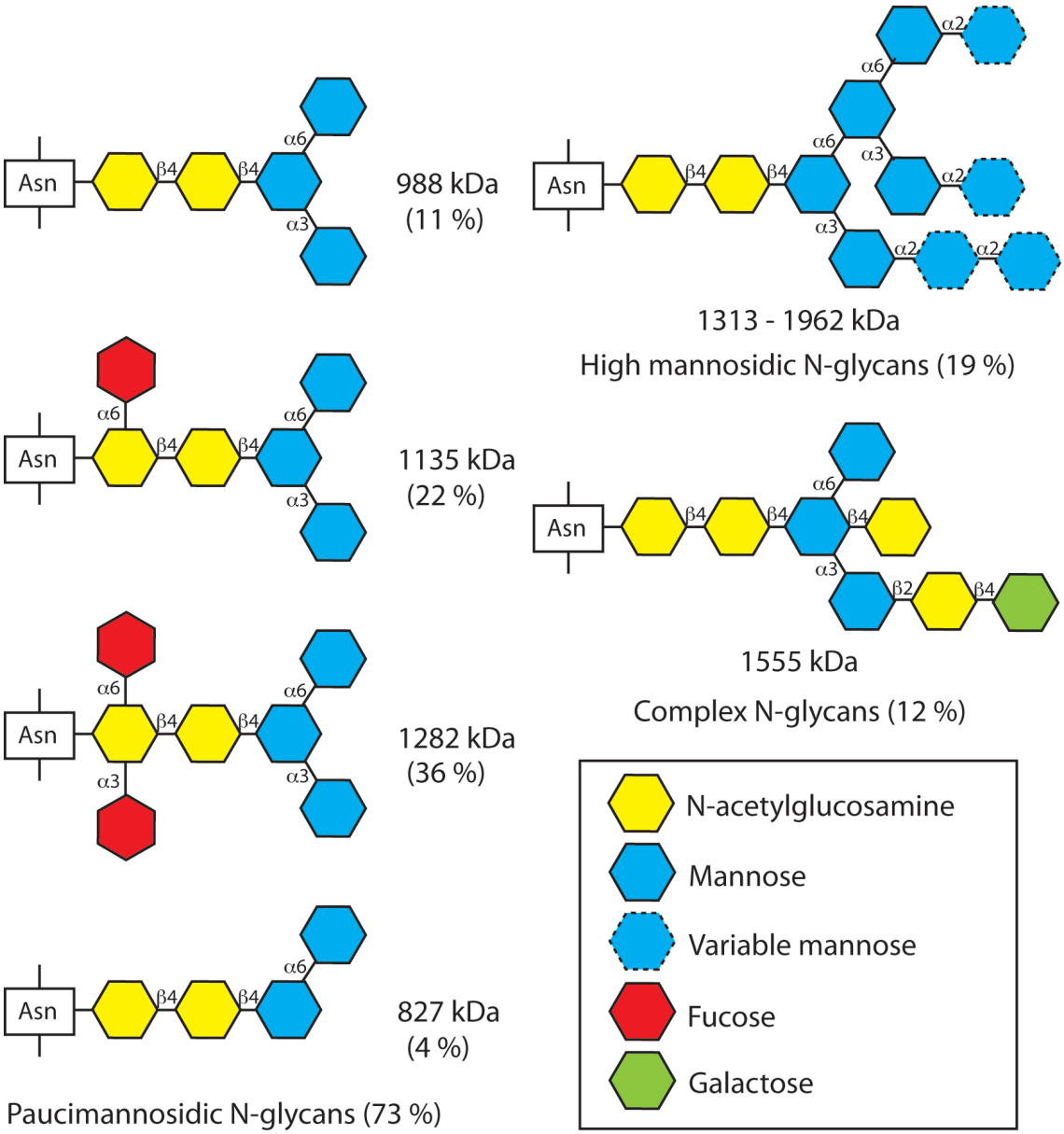



**Figure 2.9 Insect cell glycosylation**

A schematic representation of N-linked glycans attached to extracellular proteins produced and secreted by Hi5 cells. This figure is based on analysis of human β1,3-*N*-acetylglucosaminyltransferase expressed in Hi5 cells (Dojima *et al.*, 2009). It is important to note that the N-linked glycosylation pattern of proteins expressed in Hi5 cells is different for different proteins (e.g. Hsu *et al.*, 1997) and therefore the proportions and types of complex, high mannose and paucimannose glycans shown are intended to be illustrative only.

### 2.6.1 Glycosylation site prediction

zRET-ECD has 12 potential N-linked glycosylation sites bearing the NXS/T consensus, where X can't be a proline (Gavel and Heijne, 1990). Ten of these are predicted to be *bona fide* glycosylation sites using the NetNGlyc server (Table 2.4) (Gupta *et al.*, 2004). Insect cell glycosylation produces a heterogeneous mix of sugar moieties (Figure 2.9), which varies dependent on the protein being expressed. The commonly dominant species is a five-sugar paucimannose group, which may be further modified by single or double fucosylation on the innermost N-acetylglucosamine (NAG) (Dojima *et al.*, 2009). Each of these groups is between 1–1.3 kDa, adding an additional 10 – 13 kDa to the zRET-ECD protein.

| Domain | Position | Sequence | Potential | Jury agreement | N-Glyc result |
|---|---|---|---|---|---|
| CLD1 | **85** | **NKTL** | 0.7524 | (9/9) | **+++** |
| CLD1 | **137** | **NDTA** | 0.5549 | (7/9) | **+** |
| CLD2 | **185** | **NYTI** | 0.7689 | (9/9) | **+++** |
| CLD3 | **259** | **NGTD** | 0.7864 | (9/9) | **+++** |
| CLD3 | **308** | **NDSW** | 0.5710 | (7/9) | **+** |
| CLD3 | **346** | **NLSV** | 0.7008 | (9/9) | **++** |
| CLD3 | **362** | **NDTT** | 0.5100 | (6/9) | **+** |
| CLD4 | **378** | **NVTV** | 0.7172 | (9/9) | **++** |
| CLD4 | **390** | **NVTY** | 0.6464 | (9/9) | **++** |
| CLD4 | **433** | **NITA** | 0.4972 | (4/9) | **-** |
| CRD | **461** | **NDTK** | 0.6793 | (9/9) | **++** |
| CRD | **572** | **NISI** | 0.3132 | (9/9) | **---** |

**Table 2.4 Predicted N-linked glycosylation sites in zRET-ECD**

Analysis of all 12 NXS/T N-linked glycosylation consensus sites in zRET-ECD using the NetNGlyc server in ExPASY, which assesses whether the site is likely to be glycosylated on the basis of the primary amino acid sequence. Position = sequence position of the asparagine under consideration. Potential = the likelihood of glycosylation based on the averaged probability results of nine neural networks. A predicted site has a potential of > 0.5. Jury agreement = the number of the neural networks that support the prediction. N-Glyc result = indicates the prediction and the confidence level of the prediction: '+' = predicted glycosylated; '-' = predicted not glycosylated. The number of +/- indicates the confidence of the prediction, on a scale of 1–3 +/-.

### 2.6.2 Enzymatic deglycosylation

Since the stepwise removal of all 12 potential glycosylation sites would be a prolonged and slow process, as it is unknown which sites are required for protein folding and processing, the first strategy adopted for reducing the sugar content was enzymatic deglycosylation. I trialled three different deglycosylases: Peptide N-glycosidase F (PNGase F; produced recombinantly by Phillip Knowles), endoglycosidase F1 (Endo-F1; initially from Calbiochem, then produced recombinantly from a construct kindly provided by Prof. Yvonne Jones) and endoglycosidase H (Endo-H; Sigma Aldrich). PNGase F cleaves between the asparagine residue and the innermost N-acetylglucoamine (NAG), except when $\alpha$(1–3) fucosylated, removing the entire sugar moiety. Endo-F1 and Endo-H cleave between the first and second NAG moieties, leaving a single NAG attached to the asparagine residue. Endo-H does not remove sugars that have been processed by the Golgi apparatus, where the conversion from high mannosidic to paucimannosidic and then complex N-glycans occurs. Therefore this enzyme should not deglycosylate the majority of sugars on the purified secreted protein, if it has been properly processed.

The enzymes were compared at different temperatures (4, 20 and 37 °C) and reaction times (1–20 hours): PNGase F and Endo-F1 both deglycosylated the zRET-ECD, whereas Endo-H, as expected, did not (Figure 2.10). The PNGase F treated protein was prone to aggregation whereas the Endo-F1 treated material at 4 °C was much better behaved, with very minimal aggregation seen by size exclusion. This enzyme was therefore used in developing a standard protocol for zRET protein deglycosylation at 4 °C for 20 hours, followed by size exclusion chromatography to remove any aggregated protein (Figure 2.10).

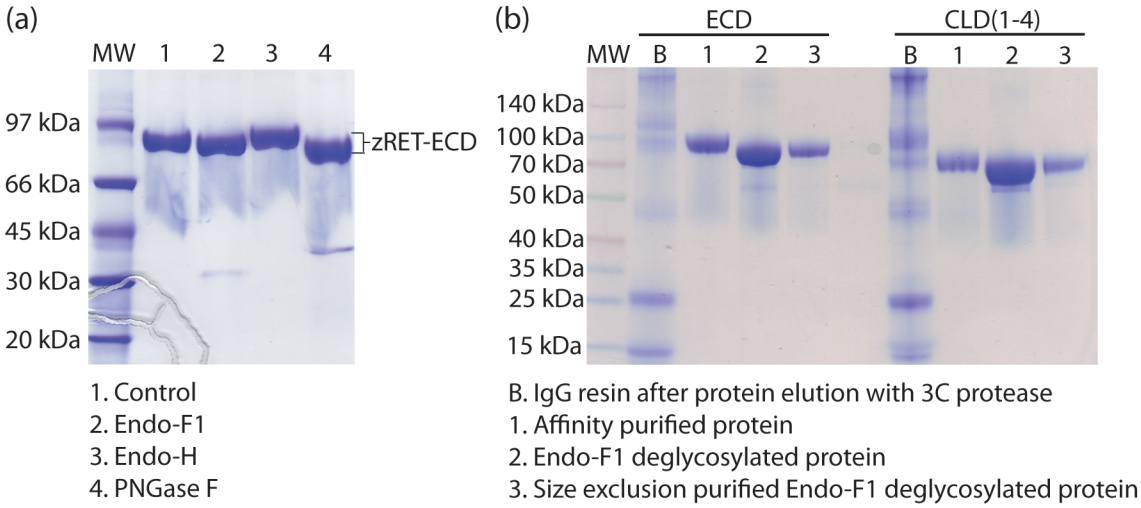



**Figure 2.10 Enzymatic deglycosylation of zRET-ECD and CLD(1–4)**

(a) Non-reducing SDS-PAGE shows results of deglycosylation of zRET-ECD at 4 °C over 3 hours with three different deglycosylases: Endoglycosidase F1 (Endo-F1), endoglycosidase H (Endo-H) and peptide N-glycosidase F (PNGase F). (b) Non-reducing SDS-PAGE showing standard Endo-F1 deglycosylation of zRET-ECD and CLD(1–4) at 4 °C over 20 hours.

Endo-F1 deglycosylation is however incomplete. Although Endo-F1 reproducibly reduces the sugar content of zRET-ECD and CLD(1–4), results from preliminary lectin column binding experiments, conducted to follow the completion of Endo-F1 deglycosylation, indicate that some mannose containing sugar groups remain on every molecule. This is corroborated by the results from SEC-MALS, which estimates the molecular weight of the Endo-F1 deglycosylated proteins to be 6–10 kDa larger than their protein molecular weight (Figure 2.11; Table 2.5). This is probably due to the low activity of Endo-F1 against fucosylated glycans. I will therefore subsequently refer to Endo-F1-treated proteins as partially-deglycosylated.

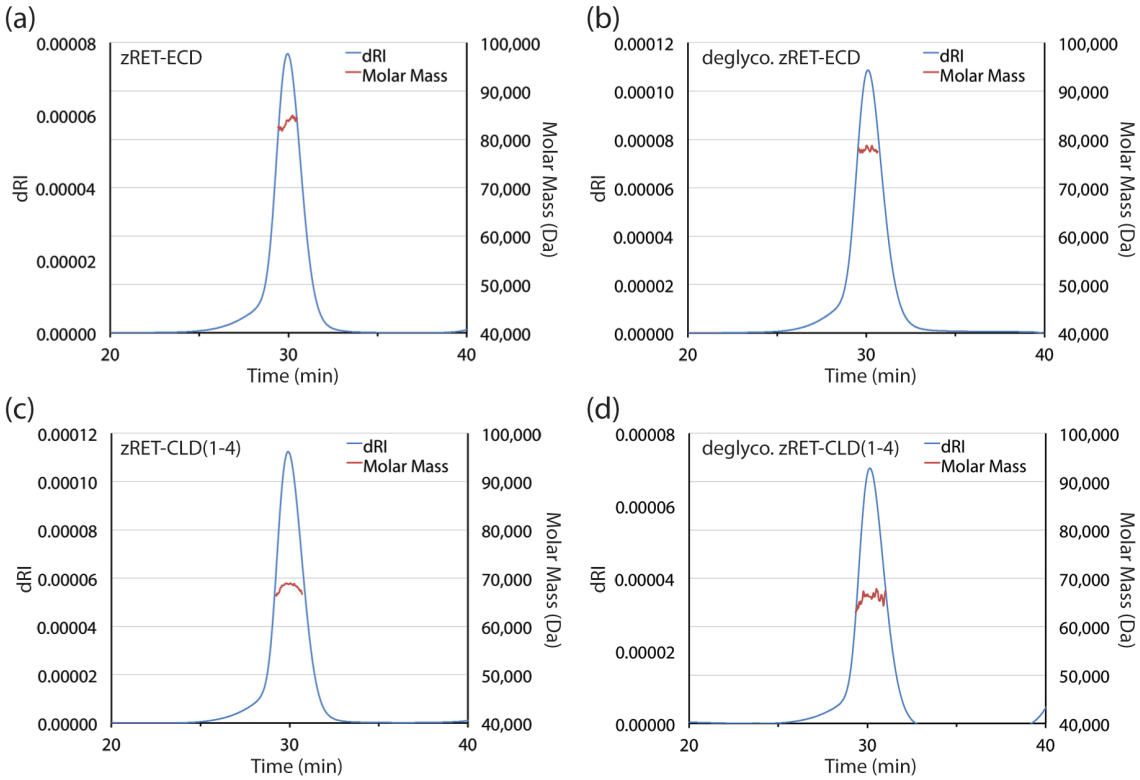



**Figure 2.11 SEC-MALS shows the zRET-ECD proteins are monomeric in solution**

SEC-MALS was carried out using a Superose 6 column for analysis of the molecular weight of previously size exclusion purified (a) zRET-ECD, (b) partially-deglycosylated (deglyco.) zRET-ECD, (c) zRET-CLD(1–4) and (d) partially-deglycosylated (deglyco.) zRET-CLD(1–4). dRI = differential refractive index.

| Protein | MALS estimate | Protein MW | Predicted Total MW |
|---|---|---|---|
| **zRET-ECD** | 83.5 kDa | 69.3 kDa | 80.8 kDa |
| **Partially-deglycosylated zRET-ECD** | 77.4 kDa | 69.3 kDa | 69.3 kDa |
| **zRET-CLD(1–4)** | 67.5 kDa | 55.7 kDa | 67.2 kDa |
| **Partially-deglycosylated zRET-CLD(1–4)** | 66.5 kDa | 55.7 kDa | 55.7 kDa |

**Table 2.5 SEC-MALS determined molecular weights of zRET-ECD and CLD(1–4)**

The table details MALS determined molecular weights from the size exclusion peak compared with the theoretical protein molecular weight and predicted total molecular weight including an estimate for the sugar content (1.15 kDa per predicted N-linked glycosylation site).

### 2.6.3 N-linked glycosylation site (NXS/T) mutagenesis

Given that enzymatic deglycosylation is incomplete, I next tested whether the zRET-ECD constructs could tolerate elimination of multiple glycosylation sites and still fold correctly in insect cells. To do this, I mutated four of the NXT/S consensus asparagines to glutamine. These four were chosen from CLD3 and 4, since this is the region with the highest number of glycosylation sites, and reducing the sugar content here might therefore be of particular help in aiding packing in a crystal. The four selected are the least well conserved glycosylation sites and may therefore be less crucial for folding (N259, N308, N378 and N433). The NetNGlyc web server did not predict that N433 is glycosylated, although the degree of certainty in this result was poor (Table 2.4).

These mutations were carried out in the context of zRET-CLD(1–4), as the most tractable crystallisation target, resulting in a construct referred to as zRET-CLD(1–4) reduced sugar (zRET-CLD(1–4)$_{redsug}$), with 7 predicted N-linked glycosylation sites.

zRET-CLD(1–4)$_{redsug}$ was expressed in insect cells and purified in the same manner as the wild type protein (section 2.3.5). The yield was around 1–1.5 mg protein per litre of culture, which is considerably lower than the wild type, but still good enough for crystallisation experiments. There was no noticeable difference in protein stability or propensity to aggregate compared to the wild type protein, as illustrated by the symmetric peak in size exclusion chromatography (Figure 2.12a). The anion exchange elution profile was slightly different to the wild type protein (see Figure 2.5), with three overlapping peaks rather than two, although as before one peak dominates (Figure 2.12b). No difference in protein from the three anion exchange peaks could be observed by non-reducing SDS-PAGE (Figure 2.12c). It is assumed the difference in the protein between the peaks is to do with the different sugar species present, although this is speculative.

This protein crystallised, as described later in this chapter (section 2.9), validating this approach for reducing sugar content in highly glycosylated regions of the target protein and thereby improving the chances of crystallisation.

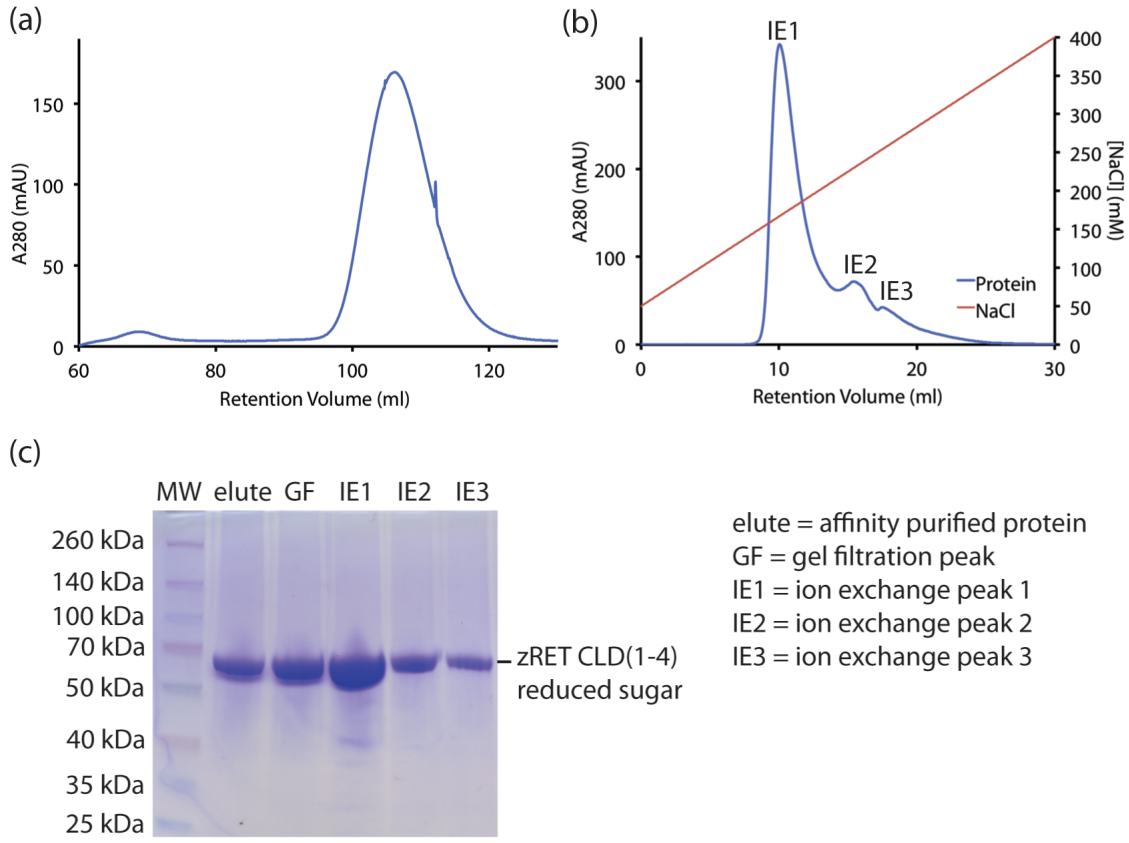



**Figure 2.12 zRET-CLD(1–4)$_{redsug}$ purification**

(a) Size exclusion chromatography of affinity purified zRET-CLD(1–4)$_{redsug}$ using a Superdex 200 column. (b) Anion exchange chromography of size exclusion purified zRET-CLD(1–4)$_{redsug}$ on a HiTrap Q column. (c) Non-reducing SDS-PAGE of protein samples from various stages of purification.

## 2.7 Defining the RET-ECD architecture by SAXS

### 2.7.1 High quality SAXS data collection at SOLEIL using online HPLC purification

In order to limit the observed effects of zRET-ECD protein aggregation on the previous SAXS data, I carried out my subsequent SAXS experiments at the SWING beamline at Synchrotron SOLEIL using their online HPLC set-up. Glycosylated (sugar content unaltered following secretion from insect cell expression hosts) and partially-deglycosylated zRET-CLD(1–4) and ECD proteins were affinity purified, subjected to size exclusion chromatography and then concentrated to ~20 mg/ml for online HPLC SAXS analysis. Approximately 20 µl of sample was loaded onto the HPLC size exclusion column, and size exclusion was carried out with the material eluting from the

column passing directly in the data collection capillary. SAXS measurements were taken across the whole protein elution peak. The data was then analysed using FOXTROT (SOLEIL in-house software), converting the 2D scattering image on the detector to a 1D SAXS profile. The apparent Rg from each data collection across the peak was determined using the Guiner approximation. For all four protein samples the Rg remained constant across the peak implying homogeneity. The data curves corresponding to the very peak of the HPLC profile for each protein were then combined and averaged to remove noise and used for subsequent data analysis.

### 2.7.2 Analysis of SAXS sample quality

To confirm the quality of the protein in the X-ray beam the SAXS data was plotted on a Kratky plot, which shows whether a protein is folded. All four proteins follow Porod's law and display a bell-shaped curve, confirming the zRET-ECD and CLD(1–4) proteins are all folded in the X-ray beam (Kratky and Porod, 1949; Putnam *et al.*, 2007). (Kratky plots of unfolded proteins don't reach a peak but only show increasing or stable $I*s^2$ with s.)

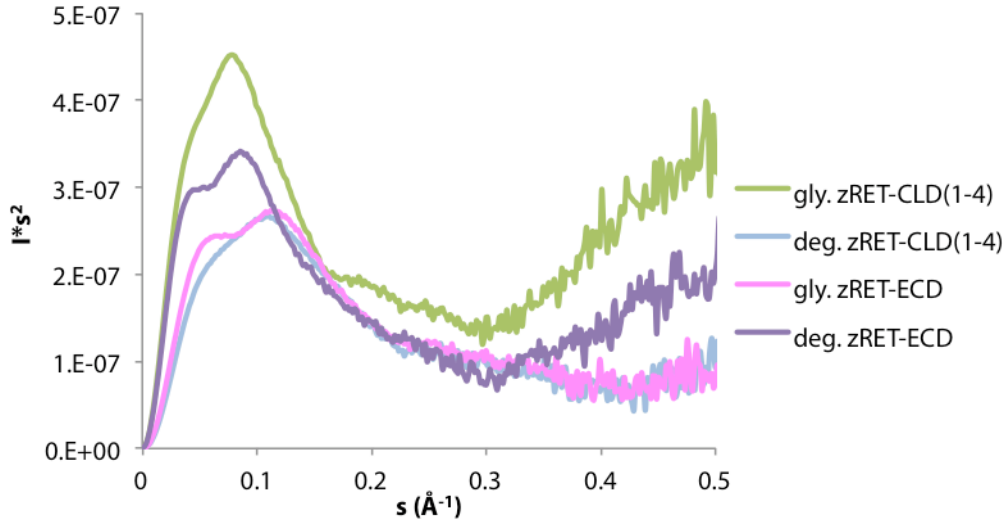



**Figure 2.13 Kratky plot shows zRET-ECD proteins are folded**

Kratky plot of HPLC coupled SAXS data generated in PRIMUS (Konarev *et al.*, 2003). gly. = glycosylated; deg. = partially-deglycosylated; s = modulus of the scattering vector; I = intensity of the scattered beam at s.

The Kratky plot can also be used to estimate the molecular weight of the sample. Determining the molecular weight from the SAXS data can reveal the oligomeric state

of the protein, and also indicate whether the sample is aggregated. This technique is generally accurate to within 10% of the actual protein molecular weight (Fischer *et al.*, 2009), although the presence of the sugars here may further decrease the accuracy of the calculation. The results of this analysis confirm that all four proteins are monomeric, in line with the SEC-MALS data (Table 2.6). Curiously the SAXS estimated molecular weight of partially-deglycosylated zRET-CLD(1–4) is larger than the glycosylated protein. The partially-deglycosylated zRET-CLD(1–4) also shows a larger Rg than the glycosylated sample (Table 2.7). This may indicate some aggregation in the partially-deglycosylated zRET-CLD(1–4) sample.

| Protein | Max S ($Å^{-1}$) | SAXS MW (kDa) | MALS MW (kDa) | Discrepancy |
|---|---|---|---|---|
| **zRET-CLD(1–4)** | 0.45 | 69.6 | 67.5 | 3% |
| **Partially-deglycosylated zRET-CLD(1–4)** | 0.4 | 77.0 | 66.5 | 16% |
| **zRET-ECD** | 0.4 | 94.9 | 83.5 | 14% |
| **Partially-deglycosylated zRET-ECD** | 0.45 | 87.5 | 77.4 | 13% |

**Table 2.6 Estimated molecular weight of zRET-ECD and CLD(1–4) from SAXS data**

Molecular weight estimate of zRET-ECD proteins from SAXS based on the area under the curve from a Kratky plot of the HPLC coupled SAXS data (SAXS MW), compared with the molecular weight estimate from MALS (MALS MW). SAXS MW calculation was carried out using SAXS MoW server (Fischer *et al.*, 2009).

### 2.7.3 HPLC coupled SAXS data processing

The SAXS data were then processed using GNOM to generate a distance distribution plot of vector distances within the protein molecules. This program generates a fit to the experimental data, based on a maximum vector length (Dmax), which is then used to generate the distance distribution plot. The results from the best fit to the experimental data for each of the zRET-ECD and CLD(1–4) protein samples are detailed in Table 2.7 and shown in Figure 2.14.

Comparing the zRET-ECD Rg of 46.0 Å from the HPLC coupled SAXS data collected at Synchrotron SOLEIL with the zRET-ECD Rg of 51.3 Å from the ESRF collected data (Table 2.2) shows a marked difference. This indicates that the setup at the Synchrotron SOLEIL beamline including the HPLC step has dramatically reduced sample aggregation.

There is only a small difference between the zRET-ECD and zRET-CLD(1–4) Rg's and maximum vector lengths (Dmax) indicating that the CRD does not particularly increase the length of the molecule. The larger Dmax value using in fitting the partially-deglycosylated zRET-ECD data produced the best fit to the data, but this larger value does not correspond to an increased particle length: The distance distribution plot of the partially-deglycosylated zRET-ECD (Figure 2.14h) shows no vectors larger than 140 Å, similar to the other three protein samples.

The distance distribution plots of the two different glycosylation states of zRET-CLD(1–4) show very similar profiles (Figure 2.14). The small difference in the amount of sugar in the sample is therefore not impacting the overall shape of the molecule. The profile is consistent with an elongated, rod-like shape. The plots of the two different glycosylation states of zRET-ECD are also very similar to each other. The profile is consistent with an elongated object but differs from the zRET-CLD(1–4) profiles by the presence of a larger second peak. This double peak profile is suggests the presence of two larger domains separated by a smaller linker.

| Protein | Rg (Å) | Rg error (Å) | Dmax (Å) | I(0) | I(0) error |
|---|---|---|---|---|---|
| **zRET-CLD(1–4)** | 44.69 | 0.067 | 144 | 5.44E-04 | 1.23E-06 |
| **Partially-deglycosylated zRET-CLD(1–4)** | 45.24 | 0.038 | 145 | 3.29E-04 | 1.56E-07 |
| **zRET-ECD** | 46.02 | 0.041 | 145 | 4.15E-04 | 2.46E-07 |
| **Partially-deglycosylated zRET-ECD** | 45.24 | 0.068 | 160 | 4.91E-04 | 5.29E-07 |

**Table 2.7 HPLC coupled SAXS data processing statistics**

Processing of HPLC coupled SAXS data with GNOM. Table details the maximum vector length (Dmax) used in fitting the data, and the real space radius of gyration (Rg) and the intensity of SAXS scattering profile extrapolated to s = 0 (I(0)) derived from the fit.

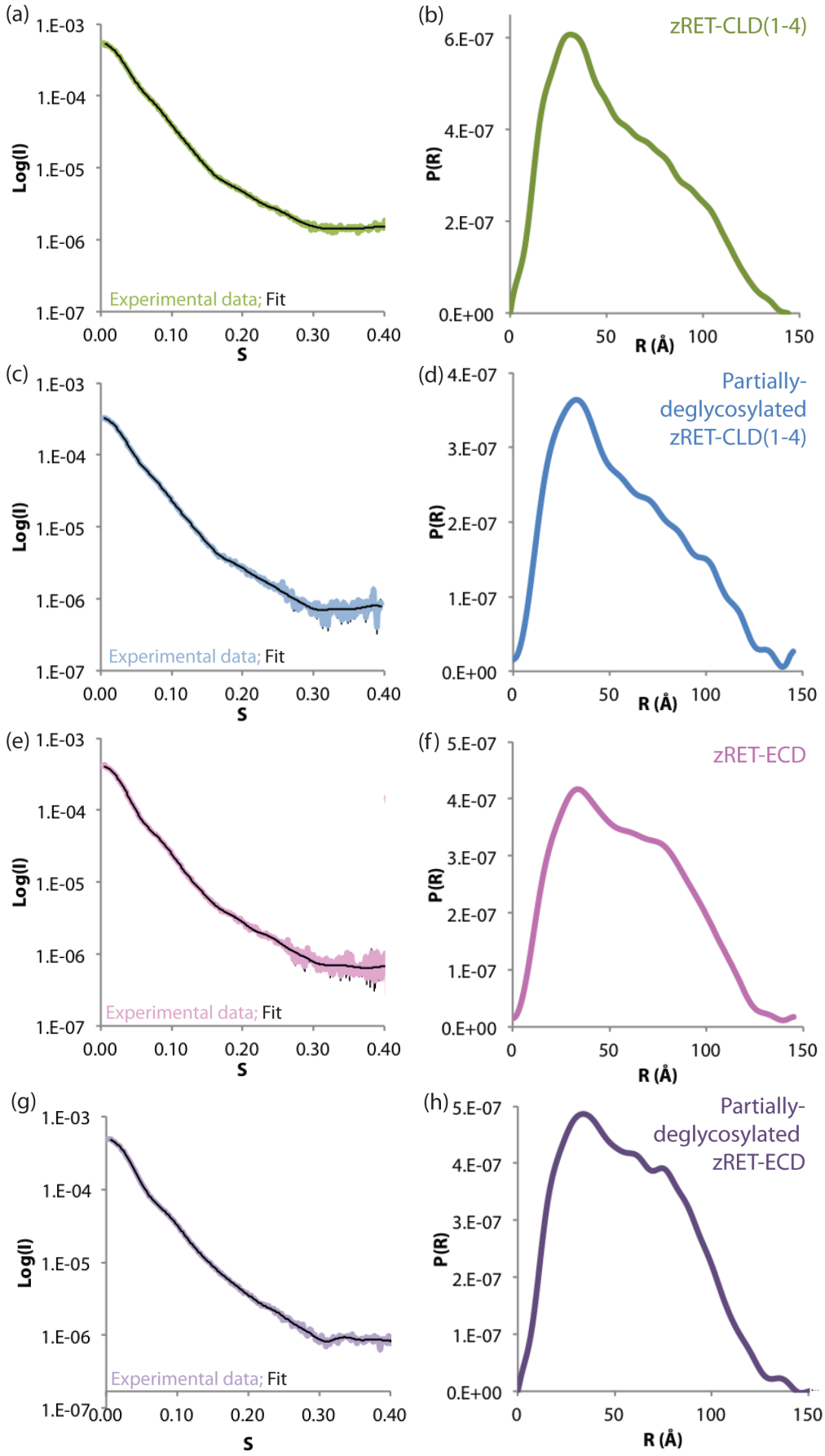

(a)
Log(I)
S
Experimental data; Fit
(b)
zRET-CLD(1-4)
P(R)
R (Å)
(c)
Log(I)
S
Experimental data; Fit
(d)
Partially-deglycosylated zRET-CLD(1-4)
P(R)
R (Å)
(e)
Log(I)
S
Experimental data; Fit
(f)
zRET-ECD
P(R)
R (Å)
(g)
Log(I)
S
Experimental data; Fit
(h)
Partially-deglycosylated zRET-ECD
P(R)
R (Å)

**Figure 2.14 zRET-CLD(1–4) and ECD SAXS data and distance distribution function**

Processing the HPLC coupled SAXS data of fully glycosylated and partially-deglycosylated zRET-CLD(1–4) and ECD with GNOM to produce distance distribution plots. (a), (c), (e) & (g) Experimental SAXS profiles compared with GNOM processing fit. s ($Å^{-1}$) = modulus of the scattering vector, I = intensity of the scattered beam at s. (b), (d), (f) & (h) Distance distribution plots derived from fit to the SAXS data. R (Å) = vector distances within the sample; P(R) = pair distance distribution function.

### 2.7.4 *Ab initio* models of zRET-ECD and CLD(1–4) from HPLC-coupled SAXS data

To gain a better understanding of the shape of the molecules the fitted data curves were used to generate *ab initio* models of the protein samples. Several models of each protein, consistent with the data, were generated in DAMMIF (Franke and Svergun, 2009), overlaid, and a final model based on the consistent features of the overlaid models was generated using DAMAVER (Volkov and Svergun, 2003) (Table 2.8).

The final *ab initio* models of the zRET-CLD(1–4) and zRET-ECD proteins show similar characteristics (Figure 2.15). As expected from the distance distribution functions, all four models show an elongated shape with an end-to-end length of 140 Å. The models also are all 'L-shaped' with a consistent bend of around 100°.

The zRET-CLD(1–4) *ab initio* models are very similar to each other, showing the same twists and bends in the long arm of the 'L-shape'. The partially-deglycosylated zRET-CLD(1–4) data indicated the sample was apparently larger than the glycosylated zRET-CLD(1–4) (section 2.7.2), which is reflected in the larger overall volume of the model. Therefore, it is difficult to attribute any differences in the two models to the differences in sugar content.

The zRET-ECD *ab initio* models show some similarity in how they differ from the zRET-CLD(1–4) models. They both have a different arrangement at the top of the long arm of the L-shape to the zRET-CLD(1–4) models, creating more of a 'C-shape'. They also both show a larger short arm of the 'L-shape'. However there are some obvious differences between the two zRET-ECD models. The short arm of the 'L-shape' is longer in the glycosylated model, and more globular in the partially-deglycosylated model.

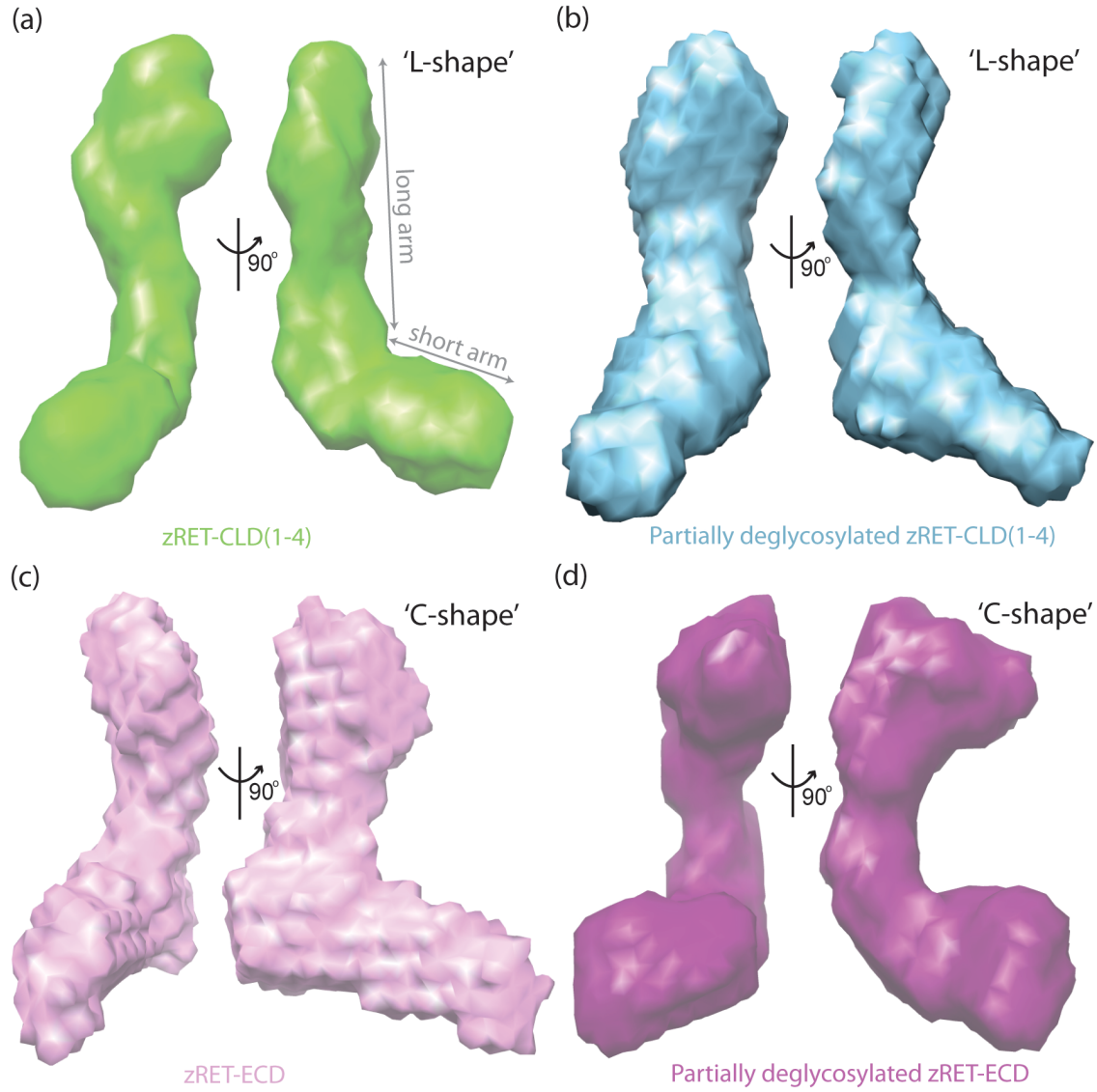



**Figure 2.15 *Ab initio* SAXS models of zRET-CLD(1–4) and ECD**

Final SAXS *ab initio* models. (a) Glycosylated zRET-CLD(1–4). (b) Endo-F1 partially-deglycosylated zRET-CLD(1–4). (c) Glycosylated zRET- ECD. (d) Endo-F1 partially-deglycosylated zRET-ECD.

| Protein | NSD mean | NSD deviation |
|---|---|---|
| **zRET-CLD(1–4)** | 0.939 | 0.031 |
| **Partially-deglycosylated zRET-CLD(1–4)** | 0.941 | 0.064 |
| **zRET-ECD** | 1.106 | 0.066 |
| **Partially-deglycosylated zRET-ECD** | 1.067 | 0.052 |

**Table 2.8 zRET-ECD and CLD(1–4) SAXS *ab initio* model generation statistics**

Generation of *ab initio* models with DAMMIF. Table details the mean correspondence (normalised spatial discrepancy, NSD) between the 10–50 generated models for each protein. These models are then overlaid and the common features are used in generation of a final model.

### 2.7.5 zRET CLDs structure prediction

In order to build a model of the arrangement of the domains within zRET-ECD based on the SAXS data, I needed the best possible structural models for each of the sub-domains. Since there are no crystal structures for any of the zRET-ECD sub-domains I had to rely on structure prediction models. The CLDs could be modelled with reasonable confidence based on cadherin structures, since they align well to consensus cadherin structural elements (Figure 1.7). The CRD has no known structural homologs and could not be modelled.

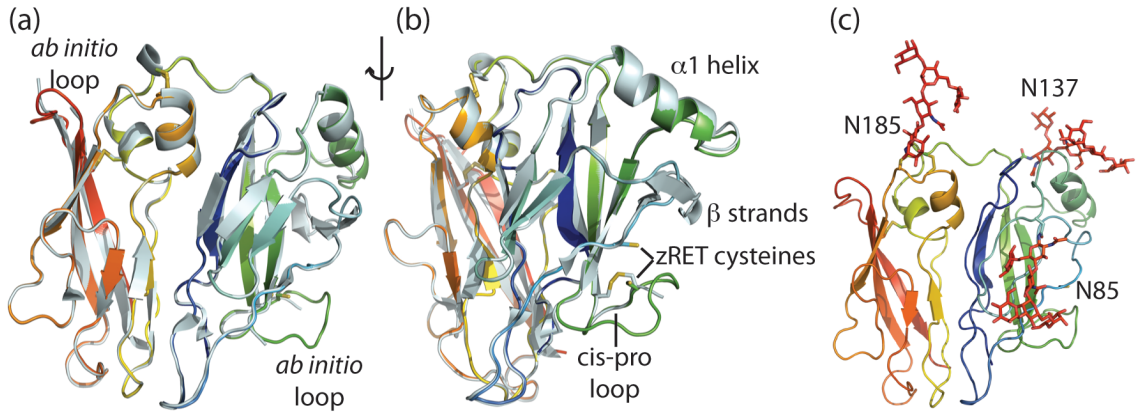



**Figure 2.16 Structure prediction model of zRET-CLD1–2**

A model of zRET-CLD(1–2) was generated using PHYRE2 intensive mode (Kelley and Sternberg, 2009) based on hRET-CLD(1–2) crystal structure (PDB 2X2U). Indicated loops were built *ab initio* using 'POING', part of the PHYRE2 server. Fucosylated paucimannose groups were added using the GLYCAM server (Woods Group, 2005-2012). (a) & (b) Ribbon representation of zRET-CLD(1–2) model, rainbow colouring, overlaid on hRET-CLD(1–2) structure, pale blue. (c) Glycosylated zRET-CLD(1–2) model with the sugars (red) and sugar-site asparagines shown as sticks.

#### *2.7.5.1 Structure prediction model of zRET-CLD(1–2) based on hRET-CLD(1–2)*

zRET-CLD(1–2) was modelled on the hRET-CLD(1–2) crystal structure (Figure 2.16). The amino acid sequence identity between these proteins is 35%. The consistent pattern of hydrophobic residues in the 7 β strands, which define the core of the cadherin-like fold in both CLD1 and CLD2, allows the zebrafish to be modelled with confidence based on the human structure (see Figure 2.2). The peripheral elements of the CLD1 structure that are not found in classic cadherin domains are not as well conserved

between zebrafish and human. These are the α1 helix, the short anti-parallel β strands and the cis-pro loop (annotated on Figure 2.16b). The α1 helix is predicted, by secondary structure prediction, to be present in the zebrafish, however the β-strands and cis-pro loop are not and therefore these loops were not based on the human structure.

The final model was glycosylated on the three predicted N-linked glycosylation sites, with a fucosylated paucimannose group using the GLYCAM server (Woods Group, 2005-2012). The fucosylated paucimannose group was chosen, as its molecular weight (1.15 kDa) is closest to the average N-linked glycan molecular weight in Hi5 produced proteins (Figure 2.9).

#### *2.7.5.2 Structure prediction models of zRET-CLD3 and zRET-CLD4 based on cadherin structures*

Structure prediction models of zRET-CLD3 and CLD4 were produced based on alignment to cadherin crystal structures (Figure 2.17). Different models were produced using PHYRE2 (Kelley and Sternberg, 2009) in normal and intensive mode and the best model for each was chosen based on the alignment to consensus cadherin elements shown in Figure 1.7. In particular the topology of the β strands and the location of the calcium-binding residues in CLD3 were compared carefully. A further constraint was the location of the predicted N-linked glycosylation sites. Models where these sites were buried were discarded. The predicted hydrophobic linker between CLD3 and CLD4 was not included in either model as its presence confused the alignments, resulting in models with incorrect β strand topology.

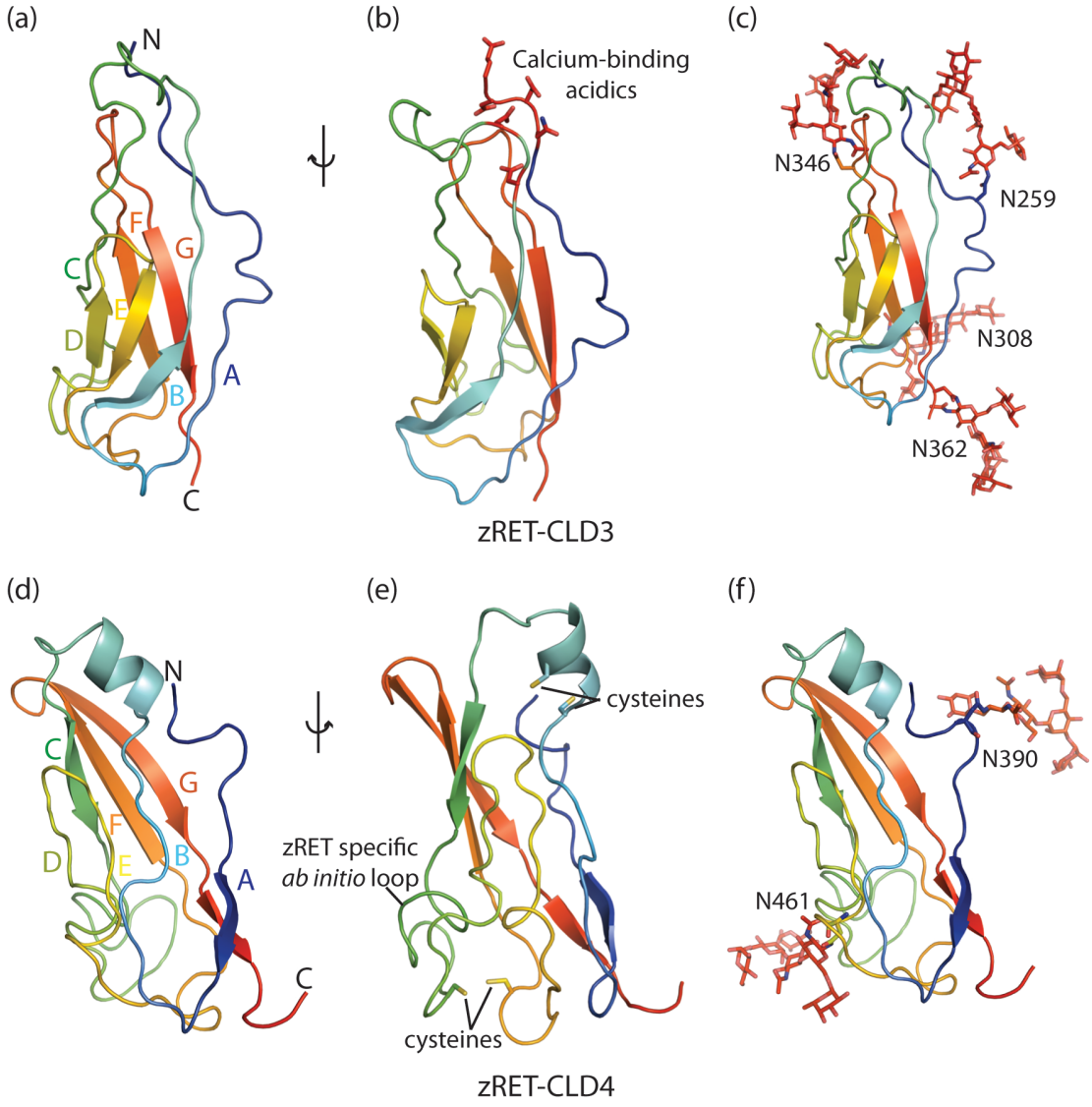



**Figure 2.17 Structure prediction models of zRET-CLD3 and CLD4**

(a), (b) & (c) Model of zRET-CLD3 based on extracellular cadherin (EC) domains from two cadherin crystal structures: C-cadherin (1L3W) and (2EE0). (a) & (b) Ribbon representation of zRET-CLD3 model, rainbow colouring. The calcium binding residues at the CLD2–3 junction are shown in stick depiction. The 7 β strands are labelled A–F, matching standard cadherin nomenclature. (c) Glycosylated zRET-CLD3 model with the GLYCAM modelled paucimannose groups (red) and linked asparagines shown as sticks. (d), (e) & (f) Model of zRET-CLD4 generated using PHYRE2 intensive mode. The model is based on 4 EC domains from three different protein structures (PDB 1OP4, 1NCI, 2X2U) and *ab initio* modelling of a zRET specific 8 residue loop. (d) & (e) Ribbon representation of zRET-CLD4 model with rainbow colouring. (f) Glycosylated zRET-CLD4 model with the GLYCAM modelled paucimannose groups (red) and linked asparagines shown as sticks.

#### 2.7.5.3 *Modelling the CLD2–CLD3 calcium bound junction based on a classical cadherin*

The calcium-binding residues in the CLD2–3 junction are conserved with cadherins (Figure 1.7) and therefore the junction could be modelled based on the arrangement between EC domains from a classical cadherin crystal structure. However, the angle between calcium-linked cadherin domains observed in crystal structures does vary by ~20–40° (angles between EC domains in C-cadherin vary from ~140–160°) and therefore the angle between CLD2–CLD3 might differ slightly from that modelled here (Figure 2.18).

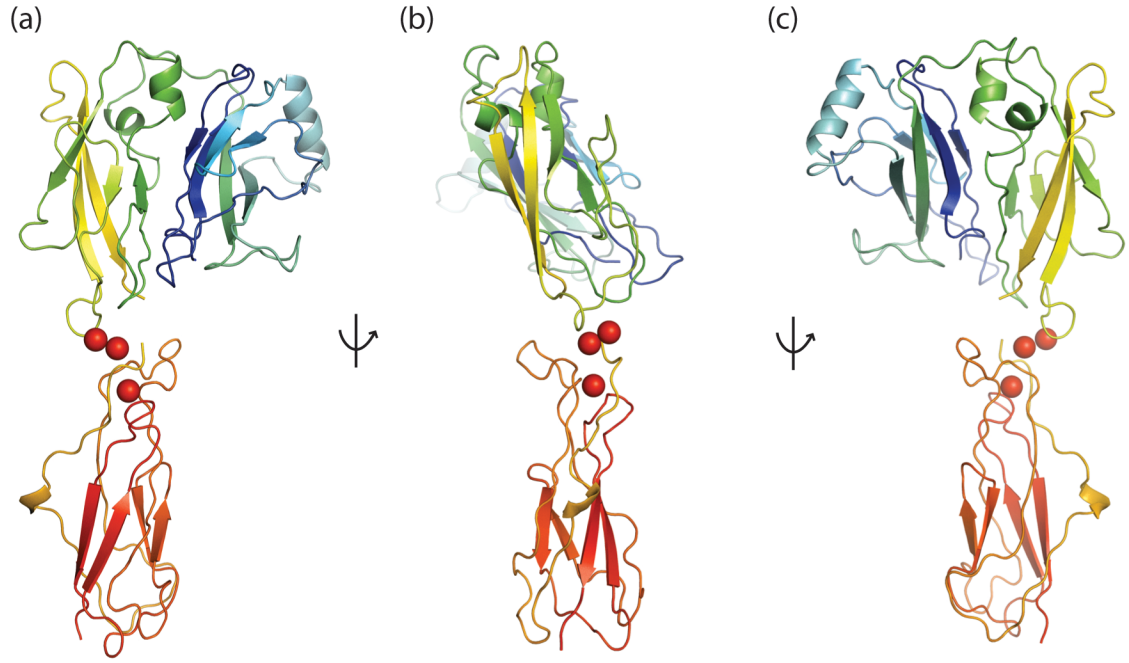



**Figure 2.18 Structure prediction model of zRET-CLD(1–3) including calciums**

zRET-CLD(1–3) structure prediction model shown in ribbon representation with rainbow colouring. Modelled calciums are shown as red spheres. CLD(2–3) junction based on alignment to EC1–EC2 domains from C-cadherin (1L3W) crystal structure, which have a 150° angle between them (Boggon *et al.*, 2002b).

### 2.7.6 SAXS-based zRET-CLD(1–4) model

In order to produce a model of zRET-CLD(1–4) based on the SAXS data I fitted the glycosylated CLD models into the zRET-CLD(1–4) *ab initio* molecular envelope using CHIMERA's fit-in-map tool (Figure 2.19a). Fitting of the domains was carried out

based on a number of assumptions. I assumed that the CLD(1–2) clam shell arrangement, with CLD1 folded back on CLD2, as in the hRET-CLD(1–2) crystal structure, would be conserved. It was also assumed that the CLD2–3 arrangement would be similar to that observed in the cadherin structures. These assumptions appeared to be reasonable and the fit of the CLD(1–3) arrangement modelled above to the *ab initio* envelopes was very good. The final assumption was that the N-terminus of the modelled CLD4 would be next to CLD3. This seems reasonable, although the linker between CLD3–4 is 20 amino acids long and therefore could span the length of CLD4 resulting in the other end of CLD4 forming the junction with CLD3. Based on these assumptions the fitted arrangement of the CLDs into the glycosylated zRET-CLD(1–4) *ab initio* envelope showed a ~100° bend angle between CLD3 and CLD4 (Figure 2.19a&c).

The goodness of fit of the resulting glycosylated CLD(1–4) model to the experimental data was determined by comparing the experimental data to a theoretical SAXS curve derived from the model. This gave a chi-square value ($\chi^2$) of 5.3 between the model and the experimental data (Figure 2.19b).

Fitting the CLD(1–2), CLD3 and CLD4 structure prediction models in the alternative orientation into the *ab initio* envelope in CHIMERA was also attempted. The fit of the CLD(1–3) structure prediction model was not good in this orientation, since it places the 100° bend angle in the *ab initio* envelope at the CLD2–CLD3 junction. The best CLD(1–4) model from fitting CLD(1–3) and CLD4 in this orientation gave a chi-square value of > 9 when compared to the experimental data. If the CLD(1–2) and CLD3 models are fitted separately in this orientation, producing a near 100° CLD2–CLD3 angle not generally observed in calcium bound EC–EC junctions, CLD(1–2) lies partly outside the *ab initio* envelope. Therefore the model based on fitting into this orientation is not as good a description of the experimental data and is less consistent with a calcium bound CLD2–CLD3 junction.

The CLD(1–4) arrangement, with a near-linear CLD2–CLD3 junction and an ~100° bend angle between CLD3 and CLD4, was corroborated by preliminary rigid body modelling of the hRET-CLD(1-2) structure and C-cadherin individual EC domains representing CLD3 and CLD4, against the experimental glycosylated zRET-CLD(1–4)

SAXS profile in SASREF (Petoukhov and Svergun, 2005). This method does not involve the use of the *ab initio* envelope. For this non-glycosylated CLD models had to be used since SASREF does not work with sugar moieties. Even though the CLD(1–2), CLD3 and CLD4 representative models were used individually, such that the CLD2–CLD3 junction angle was not enforced, the SASREF output showed a near-linear arrangement between CLD2 and CLD3 models, and an ~100° bend between CLD3 and CLD4. (Dr. Andrew Purkiss-Trew conducted this rigid-body modelling analysis.) Furthermore the goodness-of-fit of a manually built CLD(1–4) model using the glycosylated CLD structure prediction models with a linear CLD3–CLD4 arrangement was poor ($\chi^2 = 10.9$).

I used the *ab initio* envelope and experimental data from the glycosylated zRET-CLD(1–4) in modelling, since from data analysis it appeared to be the better dataset. Furthermore, since it is not known exactly how much sugar is removed in deglycosylation, I could estimate the level of glycosylation of the untreated protein with greater confidence. Nonetheless the fit of the model described here to the partially-deglycosylated zRET-CLD(1–4) dataset is actually better as measured by the chi-square value ($\chi^2 = 4.2$). Although the apparent improvement may simply be a factor of the larger experimental error ($\sigma_{exp}$) values in this dataset, since $\chi^2 = \Sigma((I_{exp}(s) - I_{model}(s))/\sigma_{exp}(s))$, which are due to the lower amount of deglycosylated zRET-CLD(1–4) sample loaded on the HPLC column for SAXS data collection compared to the glycosylated zRET-CLD(1–4) sample.

Overall the SAXS based model of the arrangement of the CLDs in CLD(1–4) (Figure 2.19c) affirms an approximately linear CLD2–CLD3 arrangement and reveals a 100° angle between CLD3 and CLD4, an unusual arrangement for cadherin domains. This arrangement is likely the result of the RET-conserved 20 amino acid, predominantly hydrophobic, linker between CLD3 and CLD4.

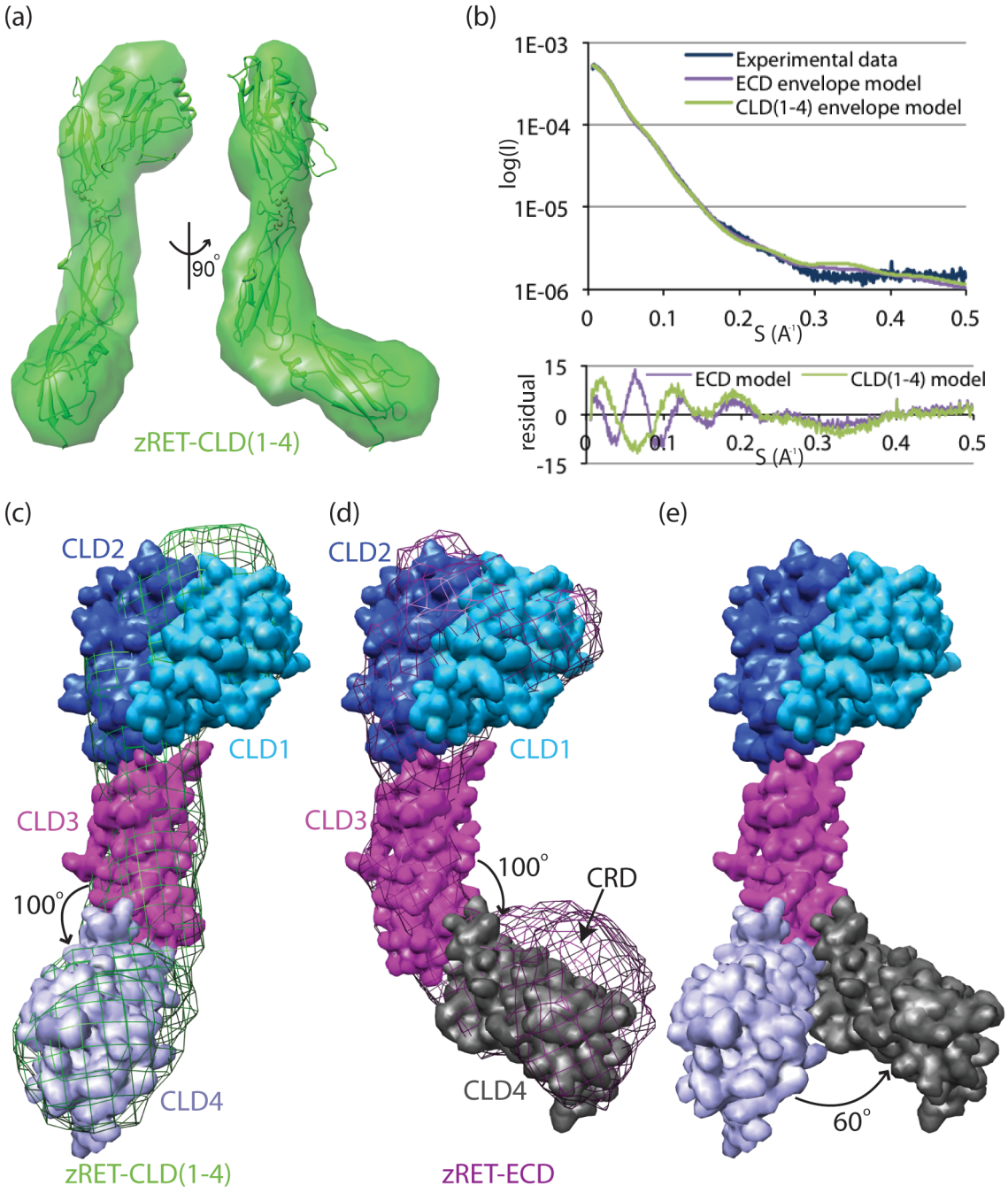



**Figure 2.19 SAXS based models of the arrangement of the CLDs in the zRET-ECD**

(a) Fitting CLD models (green ribbon representation) into zRET-CLD(1–4) SAXS *ab initio* molecular envelope (green surface). (b) Experimental SAXS data for the glycosylated zRET-CLD(1–4) compared to theoretical SAXS curves of fitted CLD models in the CLD(1–4) envelope and ECD envelope. The residual plot below shows the correspondence between the models and the experimental data. Residual = $((I_{experimental} - I_{model})/\sigma_{experimental})$. (c) Model of zRET-CLD(1–4) shown in surface view, with SAXS *ab initio* molecular envelope shown as a green mesh. (d) Fit of CLDs (surface view) into zRET-ECD molecular envelope (purple mesh). (e) Comparison between the apparent position of CLD4 in zRET-CLD(1–4) versus zRET-ECD.

### 2.7.7 SAXS-based zRET-ECD model

Since the CRD fold is unknown and therefore a structure prediction model cannot be produced with any confidence, I was restricted to simply fitting the CLDs into the zRET-ECD *ab initio* molecular envelope to define the domain arrangement. The absence of a good structural model for the CRD also meant that theoretical SAXS profiles of alternative CLD configurations could not be compared with confidence to the experimental data as I had done for interpretation of the zRET-CLD(1–4) data, since this requires a near complete model of the structure to be a useful comparison. The CLD(1–3) arrangement fitted very well into the partially-deglycosylated zRET-ECD *ab initio* envelope, leaving space for CLD4 and the CRD. CLD4 could be placed at a 100° angle to CLD3 as in the zRET-CLD(1–4) model, with the CRD folded back against CLD4 (Figure 2.19d). The CRD is the same size as a CLD in terms of amino acid sequence, and an additional CLD could almost be accommodated next to CLD4 in the zRET-ECD *ab initio* envelope. This implies that the majority of the CRD adopts a globular fold and is stably located, folded back on CLD4 in the zRET-ECD in solution.

Although the bend angle between CLD3 and CLD4 is the same in the zRET-CLD(1–4) and zRET-ECD models, there is an apparent 60° rotation about the long axis running through CLD3 between the position of CLD4 in the two models (Figure 2.19e). Since SAXS data does not specify the hand of the molecule, it was initially thought that this difference could be the result of the *ab initio* envelopes of zRET-CLD(1–4) and zRET-ECD corresponding to different hands. To test this I generated the enantiomorph of both envelopes to compare them. This revealed that the *ab initio* envelopes presented here show the greatest correspondence, and the rotation of CLD4 is not the result of the *ab initio* envelopes corresponding to different hands. Upon testing the goodness-of-fit of the CLD(1–4) model docked into the zRET-ECD *ab initio* envelope to the glycosylated zRET-CLD(1–4) SAXS data the chi-square value for this model was found to be 3.7, better than the chi-square value for the model generated by fitting in the zRET-CLD(1–4) *ab initio* envelope ($\chi^2$ = 5.2) and better than 30° rotation angles of CLD4 either side of the zRET-ECD model position. The apparent rotation of CLD4 from fitting in the *ab initio* envelopes is therefore most likely not a true difference between zRET-CLD(1–4) and zRET-ECD in solution, but a reflection of the limitations of *ab initio* models. The

location of CLD4 in the zRET-ECD model is the best description of the SAXS data based on the chi-square value.

Above I have discussed fitting into the partially-deglycosylated zRET-ECD molecular envelope as from the data analysis this appeared to be better zRET-ECD dataset (see section 2.7.4). The fit of the CLD(1–4) arrangement into the glycosylated zRET-ECD *ab initio* molecular envelope (Figure 2.15) is good but leaves unmodelled 'density' at both the CLD3–CLD4 junction and at the tip of CLD4. This could imply a more flexible arrangement of the CRD in the glycosylated protein than the partially-deglycosylated zRET-ECD *ab initio* envelope, although this cannot be proved with the available data.

## 2.8 Calcium-free RET-ECD architecture

In generating a model of the arrangement of the CLDs based on the SAXS data described in the previous section, I assumed that the CLD2–3 junction was maintained in a roughly linear arrangement by calcium coordination. To test the role of calcium in the zRET-ECD architecture, and confirm the positioning of the CLD2–3 junction within the SAXS-based model, I collected SAXS data from zRET-ECD and CLD(1–4) purified in the absence of calcium.

### 2.8.1 Calcium-free zRET-ECD protein production

zRET-ECD and CLD(1–4) were expressed in Hi5 cells, affinity purified from the media and subjected to size exclusion chromatography as usual. However all the buffers used in purification did not contain any calcium. The size exclusion profiles of protein prepared in this way show a small peak, corresponding to a molecular weight of about double the main peak, next to the main monomeric peak (Figure 2.20). This additional peak is not present in the size exclusion profile of these proteins prepared in the presence of 2 mM calcium chloride (see Figure 2.3 and Figure 2.5). The lack of calcium is therefore resulting in an increased propensity of the protein to self-associate.

In this preparation I have not tried to strip any calcium already bound to the proteins from the cell culture media. zRET-ECD produced in this way does not bind ligand/co-receptor without the addition of calcium to the buffer (data shown in Figure 4.5), and

therefore protein produced in the absence of calcium, does not have sufficient calcium bound to adopt a ligand/co-receptor binding competent arrangement.

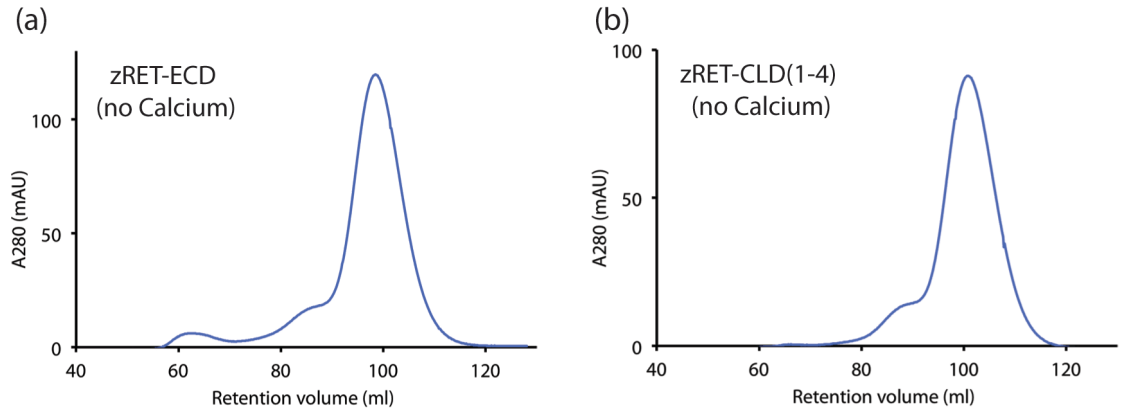



**Figure 2.20 zRET-ECD and CLD(1–4) purification in the absence of calcium**

Size exclusion profiles of (a) zRET-ECD and (b) zRET-CLD(1–4) purified without calcium in any of the purification buffers.

### 2.8.2 SAXS data collection of calcium-free zRET-ECD proteins at SOLEIL with online HPLC

SAXS data for calcium-free zRET-ECD and zRET-CLD(1–4) was collected at Synchrotron SOLEIL using the online HPLC set-up described in section 2.7. Although the protein for SAXS analysis was only taken from the monomer peak from size exclusion, the online-HPLC trace at the beamline showed the same profile as the previous size exclusion trace for both proteins. Analysis of the Rg across the HPLC peak, calculated using the Guiner approximation in FOXTROT, confirmed the presence of larger molecules in the first peak (Figure 2.21). The Rg across the main peak shows a downward trend from 42.2 Å to 40.8 Å for zRET-CLD(1–4) and 43.4 Å to 41.6 Å for zRET-ECD. This is because the two peaks are not fully separated, and therefore the data from the main peak is slightly contaminated with larger species.

This is reflected in the calculation of the molecular weight of the protein from the HPLC peak SAXS data (Table 2.9). The calculated molecular weights are still most consistent with the protein being predominantly monomeric, but they are higher than the calculated molecular weights from the SAXS data of the proteins in the presence of calcium by 5–20 kDa (see Table 2.6 for values in the presence of calcium).

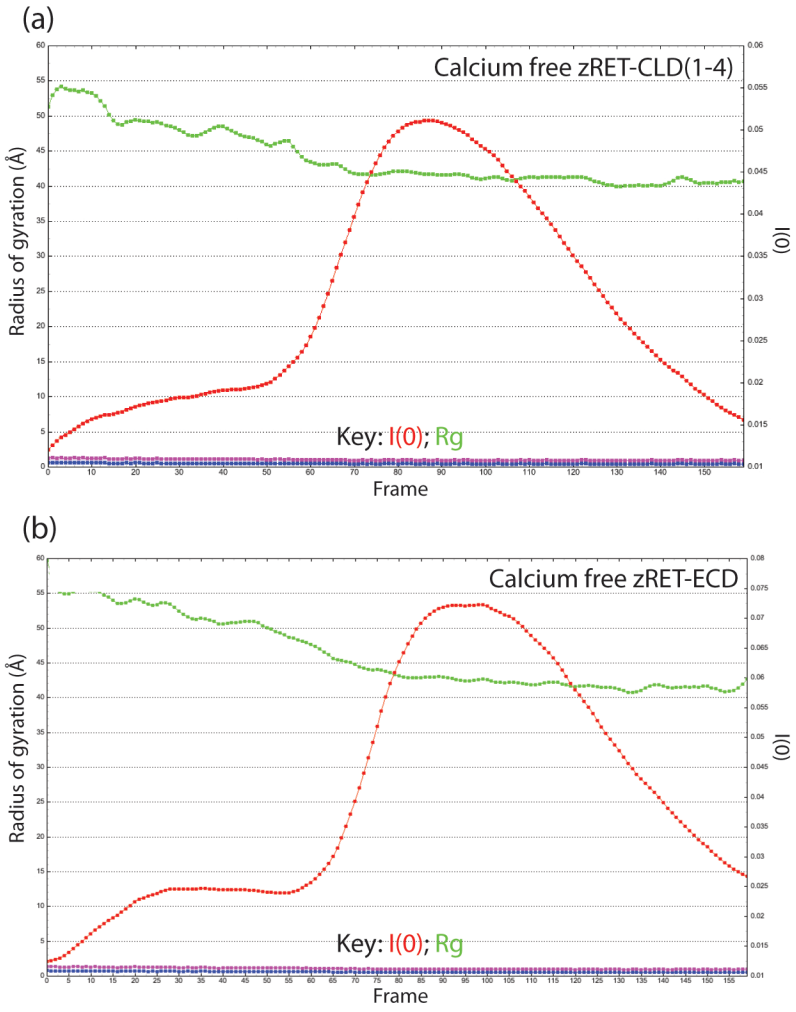



**Figure 2.21 Calcium-free zRET-CLD(1–4) and zRET-ECD SAXS data collection with online HPLC**

Plots of the Rg of the sample in the beam across the entire HPLC elution peak. The Rg and I(0) was calculated from each data collection (frame) SAXS curve using the Guiner approximation in FOXTROT. (a) Calcium-free zRET-CLD(1–4). (b) Calcium-free zRET-ECD.

| Protein | Max q | MW from SAXS (kDa) | Calculated MW (kDa) | Discrepancy |
|---|---|---|---|---|
| **Calcium-free zRET-CLD(1–4)** | 0.4 | 92.8 | 67.2 | 38% |
| **Calcium-free zRET-ECD** | 0.45 | 101.7 | 80.8 | 26% |

**Table 2.9 Estimated molecular weight from SAXS data of calcium-free zRET-ECD proteins**

Molecular weight (MW) estimate of calcium-free zRET-ECD proteins from SAXS based on the area under the curve from a Kratky plot of the HPLC coupled SAXS data. Calculation carried out using SAXS MoW server (Fischer *et al.*, 2009).

### 2.8.3 Calcium-free HPLC coupled SAXS data analysis

The SAXS data from the HPLC peak was fitted using GNOM to generate distance distribution functions for zRET-ECD and CLD(1–4) (Table 2.10 and Figure 2.22). The determined Rgs from the proteins in the absence of calcium were 2 Å lower than when calcium was present. This suggests that the overall shape of the zRET-ECD is more globular in the absence of calcium.

The distance distribution profiles for the calcium-free proteins also indicate that the overall shape is more globular in the absence of calcium. Globular proteins distance distribution plots show a symmetric peak, and the profiles for the calcium-free proteins are more symmetric than those in the presence of calcium (calcium-free: Figure 2.22; calcium-bound: Figure 2.14).

| Protein | Rg (Å) | Rg error (Å) | Dmax (Å) | I(0) | I(0) error |
|---|---|---|---|---|---|
| **Calcium-free zRET-CLD(1–4)** | 42.97 | 0.037 | 150 | 5.12E-02 | 1.46E-05 |
| **Calcium-free zRET-ECD** | 43.9 | 0.048 | 160 | 7.23E-02 | 4.66E-05 |

**Table 2.10 Calcium-free zRET-ECD proteins SAXS data processing statistics**

Processing of HPLC coupled SAXS data with GNOM. Table details the maximum vector length (Dmax) used in fitting the data, and the real space radius of gyration (Rg) and I(0) derived from the fit.

### 2.8.4 *Ab initio* models of calcium-free zRET-ECD proteins

The fitted data was then used to generate *ab initio* models of calcium-free zRET-CLD(1–4) and zRET-ECD, in order to determine a better idea of the shape of the zRET-ECD without calcium (Table 2.11 and Figure 2.22).

The final *ab initio* models of zRET-CLD(1–4) and zRET-ECD show a 'Z-shape', rather than the 'L-shape' of the models with calcium present (calcium-free: Figure 2.22; calcium-bound: Figure 2.15). The calcium-free zRET-ECD model is larger than the zRET-CLD(1–4) model both around the central portion of the Z and on one of the arms of the Z. The 'Z-shape' of these models indicates that, as expected, the linear relationship between CLD2 and CLD3 has been lost in the absence of calcium: The CLD(1–3) model can no longer fit at either end of the zRET-CLD(1–4) *ab initio*

molecular envelope. The *ab initio* envelopes therefore confirm the role of calcium in defining and rigidifying the linear CLD2–CLD3 arrangement.

The CLD(3–4) arrangement from the calcium containing SAXS model, with 100° bend angle between the domains, does fit in both envelopes, and is therefore likely maintained. However it was not possible to properly model the arrangement of the CLDs in the calcium-free zRET-CLD(1–4) *ab initio* envelope as before. If the CLD(3–4) arrangement is placed in the envelope there are two regions of spare density for the CLD(1–2): the second arm of the Z-shape, with about 100° bend angle between CLD2 and CLD3; and folded back against CLD3 in the central region of the Z-shape. However neither of these regions properly accommodate the CLD(1–2) structure, and models of both these arrangements do not match the experimental data (chi-square > 20). I therefore assume that the structure is flexible in the absence of calcium, with the CLD(1–2) in more than one position relative to CLD3.

| Protein | NSD | NSD discrepancy |
|---|---|---|
| **zRET-CLD(1–4)** | 1.000 | 0.044 |
| **zRET-ECD** | 0.928 | 0.039 |

**Table 2.11 Calcium-free zRET-ECD proteins SAXS *ab initio* model generation statistics**

Generation of *ab initio* molecular envelopes with DAMMIF & DAMAVER. Table details the mean correspondence (normalised spatial discrepancy, NSD) between the 50 generated models for each protein. These models are then overlaid and the common features are used in generation of a final model shown in Figure 2.22.

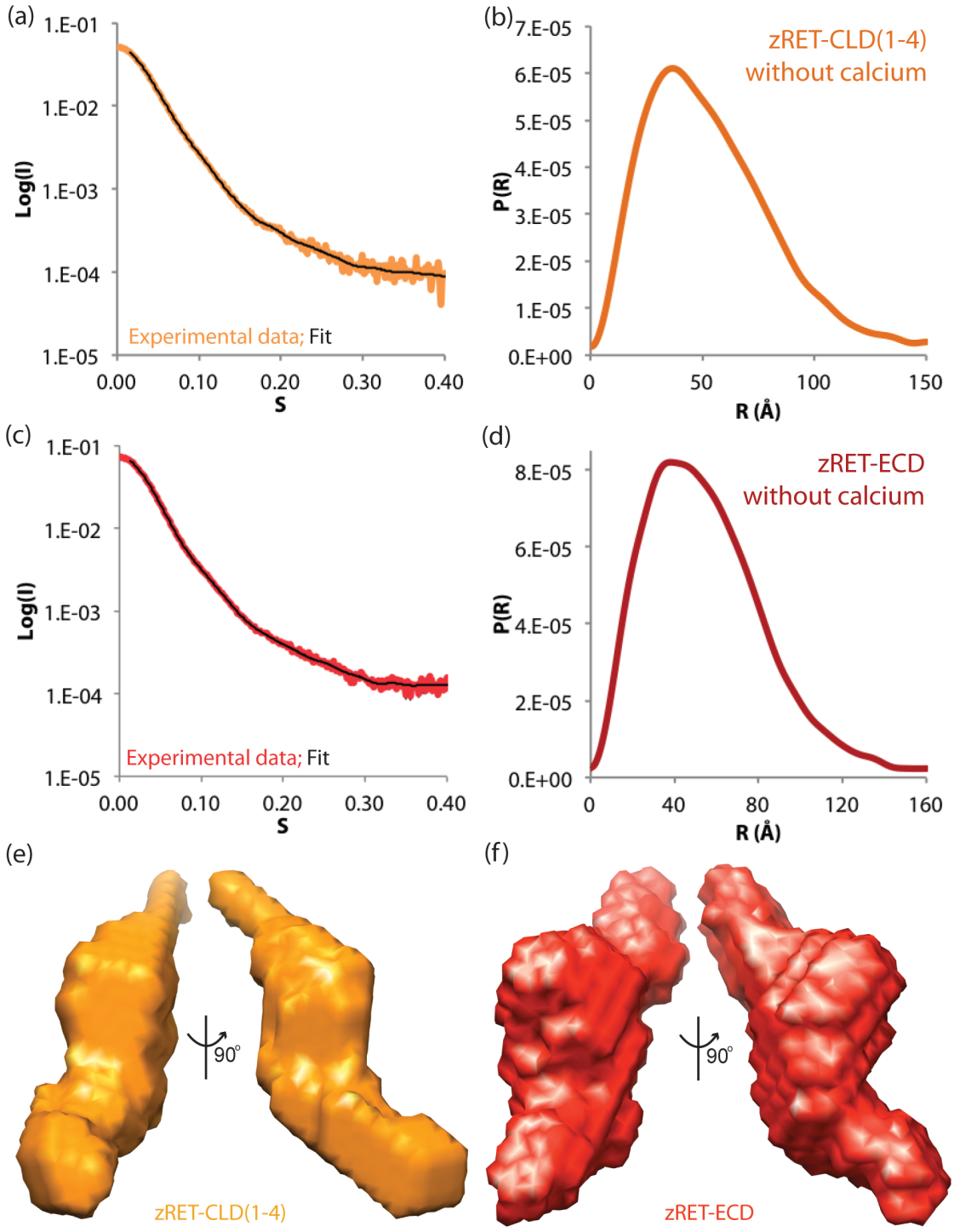



**Figure 2.22 SAXS *ab initio* models of calcium-free zRET-CLD(1–4) and ECD**

Processing the HPLC-coupled SAXS data of zRET-CLD(1–4) and zRET-ECD purified in the absence of calcium with GNOM to produce distance distribution plots for *ab initio* modelling. (a) & (c) Experimental SAXS data curves compared with GNOM processing fit. S ($Å^{-1}$) = modulus of the scattering vector, I = intensity of the scattered beam at S. (b) & (d) Distance distribution plots derived from fit to the SAXS data. R = vector distances within the sample, P(R) = pair distance distribution function. (e) *Ab initio* envelopes of calcium-free zRET-CLD(1–4) and (f) zRET-ECD.

## 2.9 zRET-CLD(1–4) crystals

Crystallisation screening of the zRET-ECD was unsuccessful. Therefore a few additional strategies were tried, including screening zRET-CLD(1–4) and the partially-deglycosylated zRET-ECD & zRET-CLD(1–4). The strategy that was successful was crystallisation of the zRET-CLD(1–4)$_{redsug}$ construct (details and purification of this construct are described in section 2.6.3). The equivalent construct of the entire zRET-ECD has not yet been produced, but given the success of the zRET-CLD(1–4)$_{redsug}$ construct this may be a fruitful strategy for future crystallisation trials of zRET-ECD.

Since the zRET-CLD(1–4)$_{redsug}$ crystals were only produced towards the end of my time in the lab, others carried out some of the work described in this section. I produced the protein, performed the initial crystallisation screening, and with the help of Dr. Andrew Purkiss-Trew, collected and processed the diffraction data from the initial crystals. Mr. Phillip Knowles then produced more protein and optimised the crystals. Diffraction data from these optimised crystals was collected and processed by Dr. Andrew Purkiss-Trew.

### 2.9.1 Crystallisation screening of zRET-CLD(1–4)$_{redsug}$ and characterisation of crystals from initial crystallisation hits

Crystallisation screening of zRET-CLD(1–4)$_{redsug}$ was carried out using material after size exclusion chromatography and after both size exclusion chromatography and anion exchange chromatography, with the three peaks from anion exchange screened separately. Crystals were produced in three different conditions, detailed in Table 2.12.

Several crystals from each of the three conditions were harvested, cryoprotected and flash frozen in liquid nitrogen. Crystal X-ray diffraction screening was carried out at Diamond Light Source I-24. The smaller crystals from the sodium formate and the MMT buffer/PEG 1500 condition were not promising. No diffraction was seen for any of the sodium formate crystals and only a few spots around the beamstop were seen for the MMT buffer/PEG 1500 condition crystals.

| Protein | Kit reference | Crystallisation condition | Crystal size | Test X-ray diffraction |
|---|---|---|---|---|
| **SE & IE, 4 mg/ml** | Wizard III/IV F7 | 25% PEG 1500, 0.1 M MMT buffer pH 6.5 | 50×25×25 µm | a few spots |
| **SE & IE, 7 mg/ml** | Salt Rx C7 | 3.5 M sodium formate, 0.1 M bis tris propane pH 7 | 40×20×20 µm | no diffraction |
| **SE, 7.6 mg/ml** | JCSG+ D1 | 24% PEG 1500, 20% glycerol | 120×120 µm | 7.5 Å |

**Table 2.12 zRET-CLD(1–4)$_{redsug}$ crystals from initial screening**

Details of zRET-CLD(1–4)$_{redsug}$ crystals produced using commercial kits. Crystal X-ray diffraction screening was carried out at Diamond Light Source microfocus beamline I24. SE & IE = size exclusion and ion exchange purified protein. SE = just size exclusion purified (see Figure 2.12). Commercial kits listed: Wizard III/IV (Emerald Biosciences), Salt Rx (Hampton Research), JCSG+ (Qiagen). PEG = polyethylene glycol, MMT buffer = L-Malic acid, MES, Tris buffer.

The larger crystals, produced in the PEG 1500/glycerol condition, were better (Figure 2.23a). Diffraction to between 7.5 and 9 Å was seen. Some evidence of multiple lattices was observed. A dataset from the crystal giving the best-looking diffraction pattern was collected in order to calculate the unit cell parameters.

The dataset was processed using D*TREK, which identified the space group as P2 (or $P2_1$), with refined cell dimensions of a = 78.51 Å, b = 88.86 Å, c = 195.6 Å, $\alpha$ = 90°, $\beta$ = 92.71° and $\gamma$ = 90°. Cell content analysis in CCP4 suggested there were 4 or 5 molecules in the asymmetric unit (probability of 0.42 and 0.48 respectively). This crystal therefore has a reasonable lattice and so optimisation of this condition to generate crystals with a single lattice and to improve resolution was carried out.

### 2.9.2 Optimisation of zRET-CLD(1–4)$_{redsug}$ crystals and data collection

Phillip Knowles reproduced the PEG 1500/glycerol condition in a larger drop size (2 µl) resulting in some better diffracting, though not larger, crystals. The best diffracting dataset obtained so far from one of these zRET-CLD(1–4)$_{redsug}$ crystals was collected at the ESRF on beamline ID23. The dataset is anisotropic, with data to 3.5 Å in the best direction, and to only 6.5 Å in the other direction. Spots from a second lattice were also evident at low resolution (Figure 2.23b). Despite these limitations Dr. Andrew Purkiss attempted processing the dataset, and this analysis revealed some details about the

arrangement of the molecules in the crystal. (Data collection statistics are detailed in Table 2.13).

| Data collection | |
|---|---|
| **X-ray source** | ESRF ID23-2 |
| **Wavelength / Å** | 0.8726 |
| **Space group** | P1 |
| **Unit cell dimensions** | |
| **a, b, c / Å** | 68.44, 108.86, 110.26 |
| **α, β, γ / °** | 106.13, 98.00, 102.85 |
| **Resolution / Å** | 40.5–3.77 (3.63–3.50) |
| **Total reflections** | 69366 |
| **Unique reflections** | 36005 |
| **Redundancy** | 1.93 |
| **Completeness / %** | 98.1 (98.4) |
| **I/σ** | 3.7 (0.7) |
| **$R_{merge}$ / %** | 10.6 (50.7) |
| **Wilson B-factor / Å²** | 72.0 |

**Table 2.13 Data collection statistics for zRET-CLD(1–4)$_{redsug}$ 3.5 Å dataset.**

The data were processed in D*TREK. Indexing was carried out using spots between 3–8 Å where data from only a single lattice was apparent. The spacegroup was found to be P1 with cell dimensions of a = 68.44 Å, b = 108.86 Å, c = 110.26 Å, α = 106.13°, β = 98.00° and γ =102.85°. Cell content analysis suggested there were 4 zRET-CLD(1–4)$_{redsug}$ molecules in the asymmetric unit.

Patterson analysis of the processed data with MOLREP (Vagin and Teplyakov, 1997) showed a weak pseudo-translation peak half way along the z-axis (Figure 2.23c). The peak height at (x = 0, y = 0, z = 0.5) is 21.24, about a third of the height of the origin peak, which is 63.82, but significantly higher than the background peaks all with a height of < 3.25. There is also some evidence of two-fold non-crystallographic symmetry from the self-rotation function, with a noisy peak of 6.11 at chi-square = 180° (Figure 2.23d), which is half the height of the origin peak (12.89) and above the background peaks of < 4.5. These findings suggest the four zRET-CLD(1–4) molecules are found in pairs related by two-fold symmetry with each pair related to the next by translational pseudosymmetry in the crystal. The presence of translational pseudosymmetry may indicate that the unit cell was originally half the apparent size

along the z-axis, but the lattice was disrupted during freezing of the crystal. Optimisation of the cryoprotection may therefore be necessary.

Optimisation of the crystallisation condition to produce better diffracting, single lattice crystals, to allow the zRET-CLD(1–4) structure to be solved is ongoing. It is hoped that with data to a reasonable resolution (3–3.5 Å or better) the structure will be solved by molecular replacement using the hRET-CLD(1–2) structure as the search model. If this is unsuccessful the calcium binding site will be exploited for heavy atom binding to determine the phases, analogous to the method used for solving the hRET-CLD(1–2) structure with a samarium derivative (Kjaer *et al.*, 2010).

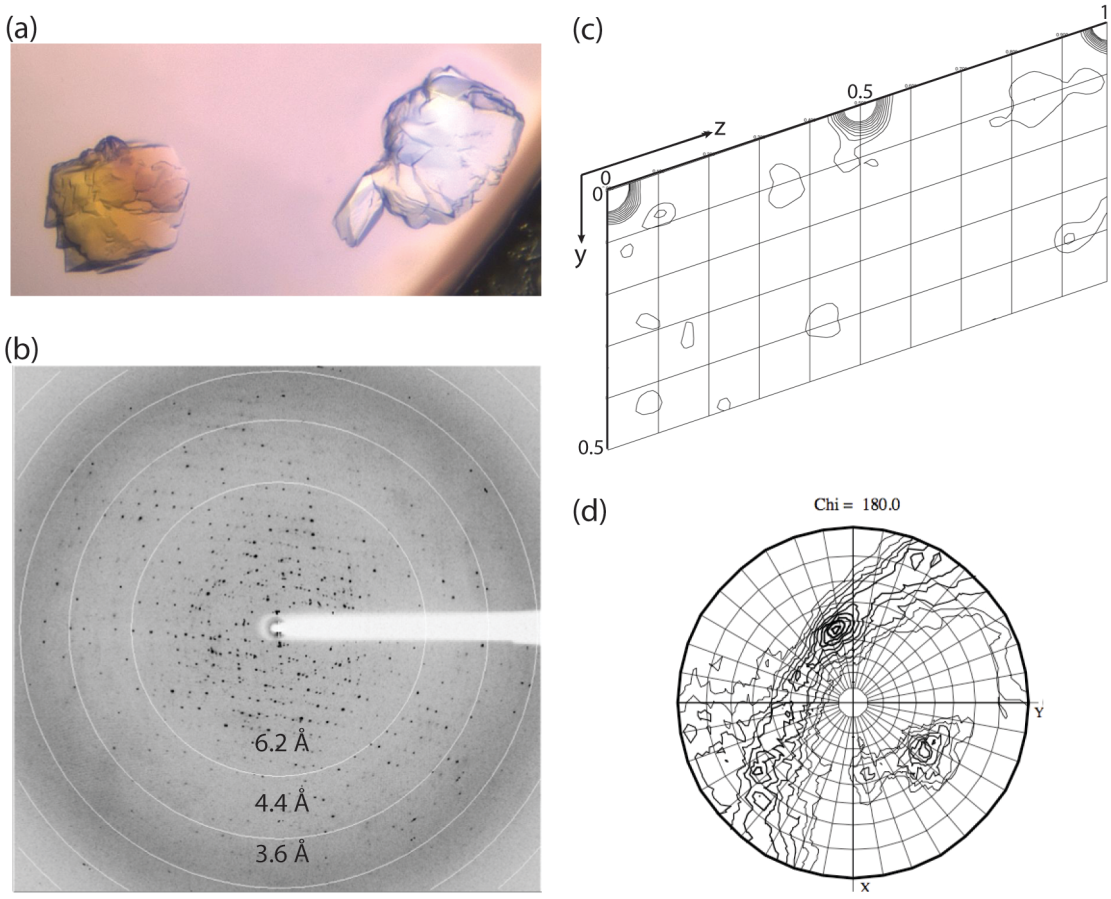



**Figure 2.23 zRET-CLD(1–4)$_{redsug}$ crystals, diffraction and patterson analysis**

(a) 120×120 µm zRET-CLD(1–4)$_{redsug}$ crystals grown in 24% PEG 1500 and 20% glycerol. (b) Best direction diffraction image from the anisotropic zRET-CLD(1–4)$_{redsug}$ dataset. (c) X-axis patterson image, showing a peak at half the unit cell length along the z-axis. (d) Self rotation function at chi = 180°.

## 2.10 Probing the structure of the CRD

The zRET-ECD SAXS model indicated that the CRD is packed against the CLD4. This suggested that producing a construct containing both CLD4 and CRD together may be soluble compared to producing individual domains and could therefore be a tractable target for crystallisation. Given the CRD's known involvement in binding ligand/co-receptor, as well as being the site of a number of cancer-causing mutations, a structural understanding of this completely structurally unknown domain would be particularly valuable.

### 2.10.1 CRD sequence analysis and secondary structure prediction

The zebrafish RET-CRD contains 14 cysteine residues and the human 16. Secondary structure prediction suggests that the zRET-CRD may be 21% alpha helix and 9% beta strand, although the confidence of the prediction for some of the elements is low. The zRET-CRD is also predicted to be around 50% disordered (Figure 2.24).

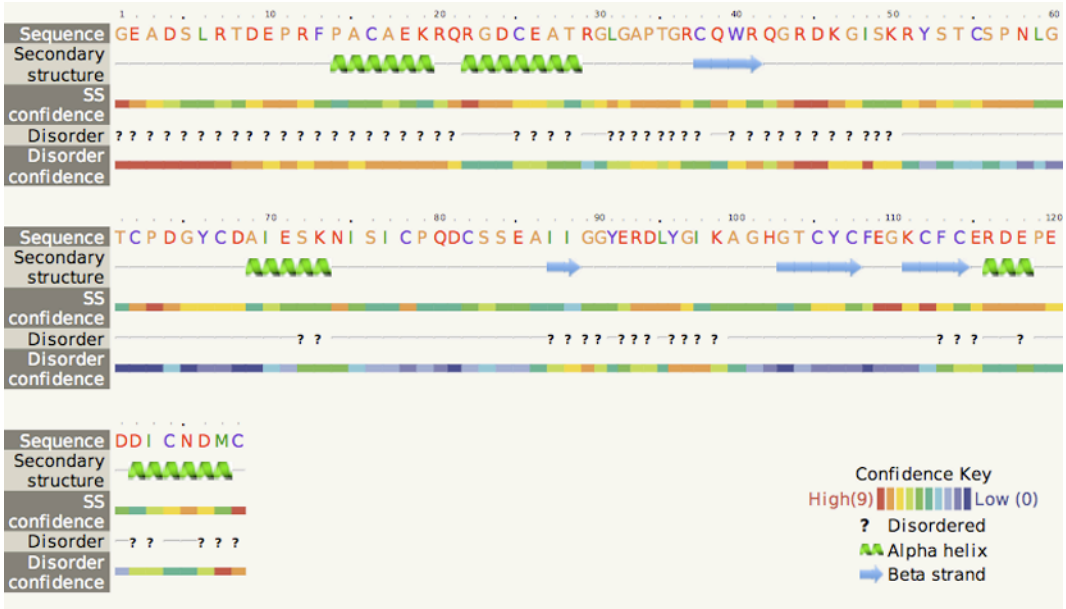



**Figure 2.24 Secondary structure prediction for zRET-CRD**

Secondary structure prediction results for zRET-CRD using PHYRE2 (Kelley and Sternberg, 2009).

### 2.10.2 zRET-CLD4CRD protein expression and purification

In order to probe the structure of this intriguing portion of the RET-ECD further I designed two zRET-CLD4CRD constructs with differing N termini. The longer construct contains residues 373–626 and the shorter residues 382–626. The ambiguity of the end of the hydrophobic CLD3–CLD4 linker and the start of CLD4 is reflected in the two constructs, since CLD4 doesn't contain the cadherin consensus PXF/Y motif, which defines the start of most cadherin domains (CLD alignment to classical cadherin domains is shown in Figure 1.7). The same pBacPAK transfer vector and expression cassette used in the zRET-ECD production, featuring the zRET native signal sequence and a C-terminal Protein A tag, was used for the CLD4CRD baculovirus production. zRET-CLD4CRD protein expression in Hi5 cells from these baculoviruses was around 1.4 mg per litre cells for both the longer and shorter constructs.

The affinity purified proteins for both the longer and shorter constructs were part monomer, part dimer. The dimer is disulphide-linked, since the amount of dimer is reduced upon addition of reducing agent to the SDS-PAGE loading buffer (Figure 2.25). The two oligomeric forms could be separated by size exclusion chromatography and ion exchange chromatography (Figure 2.26). The second (monomer) peak from ion exchange of the shorter construct has a shoulder, which contains both monomer and dimer. Due to this I decided the best construct to pursue for crystallisation was the longer one (zRET-CLD4CRD$_{373–626}$) since the purified and separated monomer and dimer would be more homogeneous.

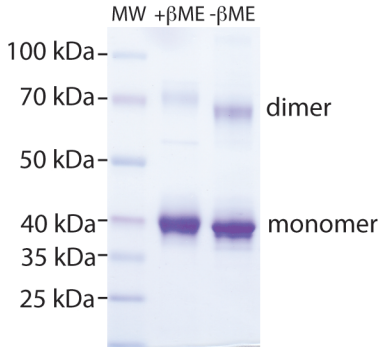



**Figure 2.25 Disulphide-linked zRET-CLD4CRD dimer**

SDS-PAGE of affinity purified zRET-CLD4CRD$_{373–626}$ with or without β-mercaptoethanol (βME) in the sample loading buffer.

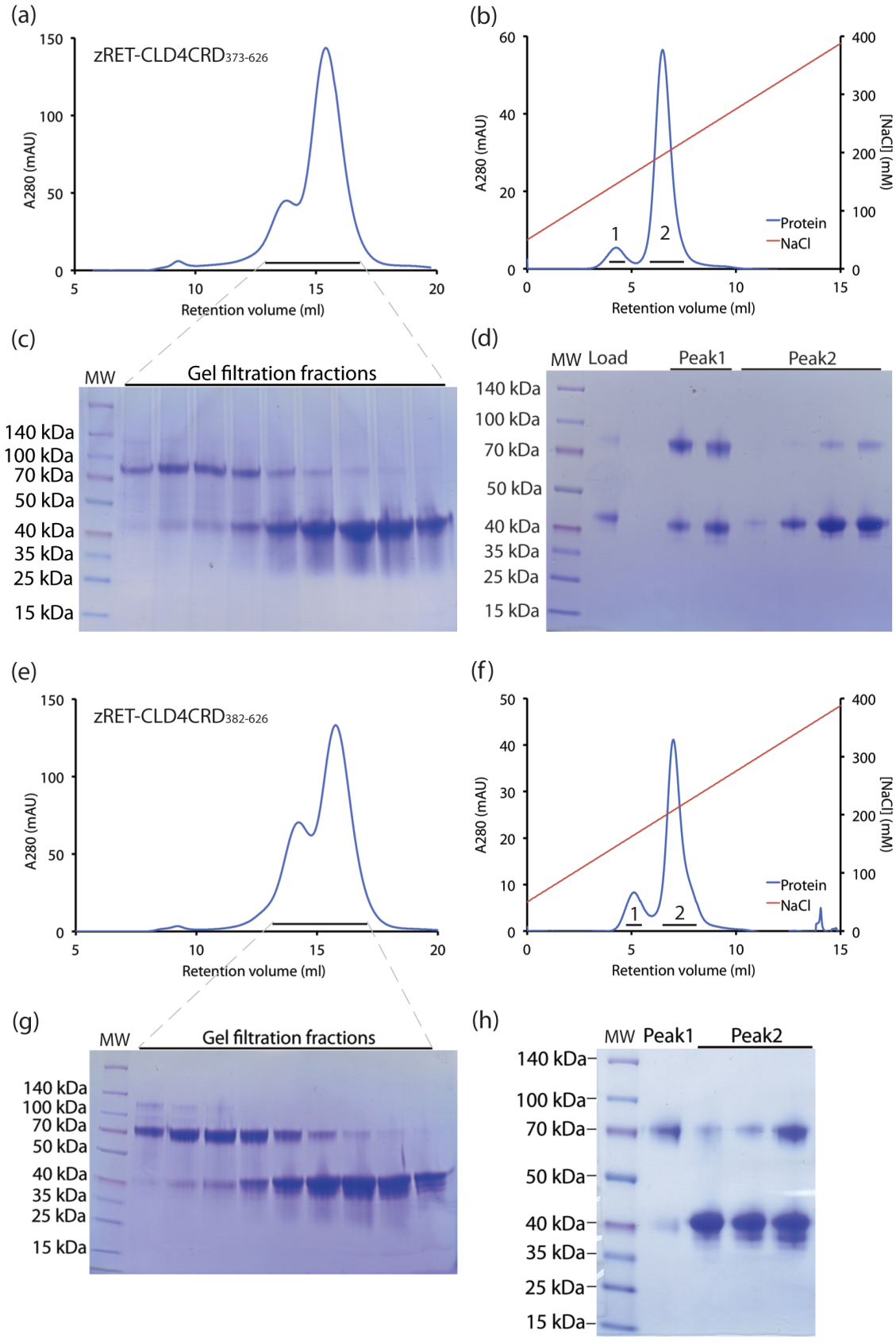

(a)
zRET-CLD4CRD373-626
A280 (mAU)
Retention volume (ml)
(b)
A280 (mAU)
[NaCl] (mM)
Protein
NaCl
1
2
Retention volume (ml)
(c)
MW
Gel filtration fractions
140 kDa
100 kDa
70 kDa
50 kDa
40 kDa
35 kDa
25 kDa
15 kDa
(d)
MW Load
Peak1
Peak2
140 kDa
100 kDa
70 kDa
50 kDa
40 kDa
35 kDa
25 kDa
15 kDa
(e)
zRET-CLD4CRD382-626
A280 (mAU)
Retention volume (ml)
(f)
A280 (mAU)
[NaCl] (mM)
Protein
NaCl
1
2
Retention volume (ml)
(g)
MW
Gel filtration fractions
140 kDa
100 kDa
70 kDa
50 kDa
40 kDa
35 kDa
25 kDa
15 kDa
(h)
MW
Peak1
Peak2
140 kDa
100 kDa
70 kDa
50 kDa
40 kDa
35 kDa
25 kDa
15 kDa

**Figure 2.26 zRET-CLD4CRD purification**

Size exclusion and anion exchange chromatography of long and short zRET-CLD4CRD proteins, showing separation of the monomeric and dimeric species. Size exclusion using Superdex 200 column of (a) zRET-CLD4CRD$_{373–626}$ and (d) zRET-CLD4CRD$_{382–626}$. Anion exchange using HiTrap Q column of (b) zRET-CLD4CRD$_{373–626}$ and (e) zRET-CLD4CRD$_{382–626}$. (c), (d), (g) and (h) Non-reducing SDS-PAGE of chromatography fractions.

### 2.10.3 The intermolecular disulphide link in the CLD4CRD dimer

The fact that the zRET-CLD4CRD proteins are expressed both as monomers and disulphide-linked dimers is curious since no disulphide-linked dimers are observed in the similarly produced whole zRET-ECD. Circular dichroism of the separated zRET-CLD4CRD$_{382–626}$ monomer and dimer showed that both species were folded and had almost identical circular dichroism spectra (Figure 2.27b). Dimerisation therefore does not affect the fold of the protein. Furthermore the ability of zRET-CLD4CRD to form disulphide-linked dimers is not due to the produced protein being unfolded.

To detect if any free cysteines were present in the purified zRET-CLD4CRD$_{373–626}$ monomer and dimer species we used Ellman's reagent. This showed that the monomer contained no reactive unpaired cysteines and analysis of the dimer showed there was approximately one unpaired cysteine per five protein molecules. These data show that the monomers have been able form all the intramolecular disulphides and this species should therefore be stable. The presence of a low level of free cysteines in the dimer is presumably because the dimer must contain two intermolecular disulphides in order for all the cysteines to be paired, and therefore in a subset of the dimers there is only one intermolecular disulphide leaving two unpaired cysteines in the dimer. (Mr. Phillip Knowles performed this Ellman's assay.)

### 2.10.4 Circular dichroism suggests the CRD has a largely unstructured fold

Circular dichroism measurements were carried out on purified zRET-ECD, zRET-CLD(1–4) and zRET-CLD4CRD$_{382–626}$ to give some indication of the secondary structure of the CRD. These measurements were carried out by Dr. Stephen Martin at the National Institute of Medical Research. As expected the zRET-CLD(1–4) profile

shows a folded, predominantly beta sheet protein. The zRET-ECD profile is fairly similar to that of CLD(1–4), again suggesting a predominantly beta sheet structure. There is a small difference between the two curves. Taking this difference and plotting it normalised to the average residue weight gives a noisy curve, from which some general interpretations about the CRD can be made (Figure 2.27). Analysis of the secondary structure content from the derived CRD curve suggests a folded domain with perhaps only a third of the domain consisting of beta sheet and alpha helical elements. Although given the high level of noise in derived CRD curve and the limitations of secondary structure prediction from circular dichroism data these values are indicative only (Table 2.14).

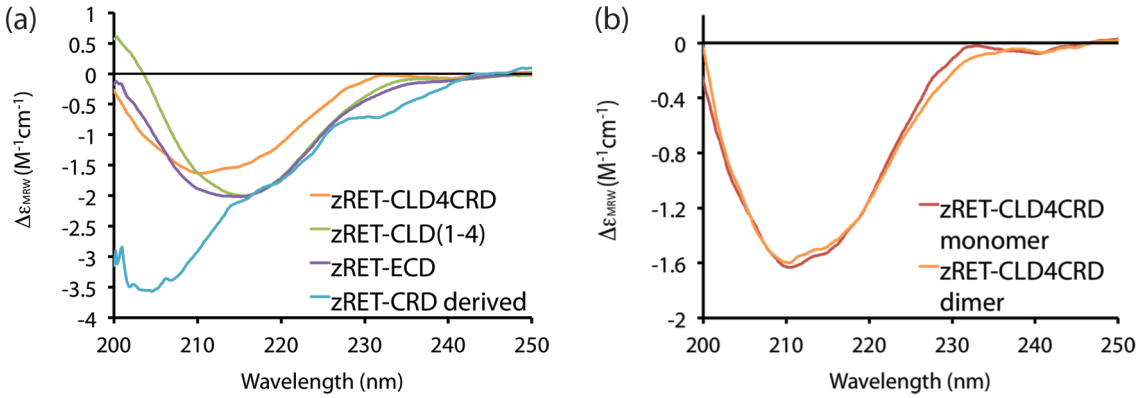



**Figure 2.27 Circular dichroism spectra of zRET-ECD proteins**

(a) Far-UV circular dichroism spectra of zRET-ECD, zRET-CLD(1–4) and zRET-CLD4CRD$_{382–626}$. The circular dichroism data is shown as the circular dichroism absorption coefficient calculated on a mean residue weight basis (ΔϵMRW). The derived zRET-CRD spectrum is calculated from the difference between the zRET-ECD and zRET-CLD(1–4) spectra. (b) Comparison between the circular dichroism spectra of size exclusion separated monomeric and dimeric zRET-CLD4CRD$_{382–626}$.

| Predicted secondary structure | CLD4CRD monomer | CLD4CRD dimer | ECD | CLD(1–4) | CRD (derived) |
|---|---|---|---|---|---|
| **Alpha helix** | 5% | 5% | 7% | 5% | 10% |
| **Beta strand** | 37% | 38% | 35% | 36% | 26% |
| **Turn** | 21% | 21% | 23% | 25% | 19% |
| **Unordered** | 36% | 35% | 36% | 34% | 46% |

**Table 2.14 Predicted secondary structure of zRET-ECD proteins based on circular dichroism spectra**

The table details predicted secondary structure from analysis of the circular dichroism spectra. The values shown are the averaged results from three different standard programs (CONTIN, SELCON and CDSSTR; Martin and Schilstra, 2008; Sreerama and Woody, 2000).

### 2.10.5 Crystallisation trials of zRET-CLD4CRD

Preliminary crystal trials of the size-exclusion purified zRET-CLD4CRD monomer and dimer species, conducted by both Mr. Phillip Knowles and I at 20 °C using a range of commercial crystallisation screening kits failed to produce protein crystals. The monomer was screened at 9 mg/ml and the dimer at 3 mg/ml. Changes to the protein buffer composition or removal of the glycosylation sites could be tried in future attempts at crystallisation, although the ability of the zRET-CLD4CRD proteins to form disulphide-linked dimers hints at flexibility within this protein.

## 2.11 Conclusions

In summary in this chapter I have presented the successful production of properly folded, highly soluble zRET-ECD, zRET-CLD(1–4) and zRET-CLD4CRD proteins in insect cells, at yields amenable to structural characterisation using X-ray crystallography and SAXS methods.

High quality SAXS data, collected for zRET-ECD and zRET-CLD(1–4) using an online HPLC purification step, revealed the arrangement of the four CLDs and the CRD in the zRET-ECD in solution. Due to conservation of the elements defining the overall architecture of the RET-ECD between lower vertebrates and mammals this arrangement is likely conserved amongst vertebrate RETs. The clamshell arrangement of CLD(1–2) previously observed in the human CLD(1–2) crystal structure, is maintained by a

conserved disulphide, and fits well into the zRET SAXS envelopes. The quasi-linear arrangement of CLD2–CLD3 is defined and rigidified by calcium binding to conserved acidic residues, confirmed by the flexibility at this junction observed by SAXS analysis of calcium-free zRET-ECD. The CLD3–CLD4 junction displays an atypical cadherin domain arrangement with a 100° bend between the two domains, presumably defined by the conserved hydrophobic linker between these domains. The CRD apparently folds back on CLD4 in the context of the whole zRET-ECD. However in the absence of CLD(1–3) the CLD4–CRD arrangement may be different, indicated by the formation of disulphide-linked *trans*-dimers when these domains are expressed alone. This model of the overall architecture and domain arrangement of the RET-ECD provided a necessary basis for interpreting an EM 3D reconstruction of the RET-ECD/GDNF/GFRα1 ternary complex presented in the next chapter.

In this chapter I have also presented a method for the crystallisation of zRET-CLD(1–4), following reduction of the sugar content of CLD3 and CLD4 by site directed mutagenesis of N-linked glycosylation sites. These crystals require optimisation to improve the resolution of the X-ray diffraction data and allow the structure to be solved. If this is successful the crystal structure will provide key insights into the structural basis for the CLD3–CLD4 bend, as well as revealing the lower vertebrate specific disulphide and loop arrangement in CLD1.

# Chapter 3. Electron microscopy reconstruction of the human RET ectodomain bound to ligand and co-receptor

## 3.1 Chapter aims

This chapter describes the production of a negative stain electron microscopy (EM) 3D reconstruction of the extracellular portion of a human RET/ligand/co-receptor ternary complex. Known or modelled structures of individual RET ternary complex components were fitted into the 3D reconstruction to give an interpretation of the arrangement of the components in the complex and generate a pseudo-atomic model. Other models were considered but a single 'best' interpretation emerged from this analysis. The overall aims of this chapter were to: (1) define the architecture of the RET-ECD/GDNF/GFRα1 ternary complex, including the relative positioning of protein subdomains and interaction regions; (2) aid the interpretation of published human/*Xenopus* chimeric RET studies as well as crosslinking and mutational investigations of the interaction between RET and ligand/co-receptor; (3) investigate any conformational changes which might occur in the RET-ECD upon ligand interaction by comparing a ligand-free form of RET-ECD with its conformation within the EM 3D reconstruction.

## 3.2 A negative stain EM 3D reconstruction of a mammalian ternary complex

### 3.2.1 Statement of contributions

The work described in this chapter was a collaborative effort primarily between myself, Dr. Svend Kjær (Cancer Research UK, London Research Institute) and Dr. Fabienne Beuron (Imperial Cancer Research, London). Dr. Svend Kjær produced the mammalian ternary complex sample. Dr. Fabienne Beuron produced the EM grids of the sample, collected the images, and did the initial sample characterisation. Under the supervision of Dr. Edward Morris, Dr. Fabienne Beuron and Dr. Paula da Fonseca, I picked

particles, generated classes, and performed angular reconstitution to generate a 3D reconstruction. Dr. Fabienne Beuron did some additional particle picking and performed the subsequent projection matching procedure to generate the final 3D reconstruction.

### 3.2.2 Protein production

In order to look at the architecture of a human RET/ligand/co-receptor complex Dr. Svend Kjær produced and assembled a recombinant ternary complex of the human RET-ECD with human GDNF (ligand) and rat GFRα1 (co-receptor), subsequently referred to as the mammalian ternary complex (mTC). Specifically, the mTC sample consisted of: the entire human RET ECD (residues 29–634, following signal sequence cleavage; hRET-ECD); the whole rat GFRα1, excepting 41 C-terminal residues containing the GPI-link site and GPI signal sequence (residues 19–427, following signal sequence cleavage; rGFRα1); and the entire mature human GDNF (residues 77–211; hGDNF). The hRET-ECD construct contained two mutations (C87R and C216S) to remove the two unpaired destabilising cysteines in hRET-CLD(1–2), in order to improve expression levels in mammalian cells, as discussed in (Kjaer *et al.*, 2010). The rGFRα1 construct contained a single asparagine to glutamine (N59Q) mutation to remove a partially occupied N-linked glycosylation site.

The hRET-ECD and rGFRα1 were expressed in stably transfected Chinese Hamster Ovary (CHO) Lec8 cells as TEV-protease cleavable Protein A fusion proteins using the method described in (Kjaer *et al.*, 2010). 2 litre batches of CHO cells were grown in suspension for 4–5 days prior to harvesting the cell supernatant. Both hRET-ECD and rGFRα1 Protein A fusion proteins were then purified from the harvested and filtered media by incubation with 0.5 ml IgG-sepharose resin overnight at 4 °C. Following washing of the resin hRET-ECD was eluted by cleavage of its Protein A tag with TEV-protease. The washed rGFRα1-*TEV*-Protein A fusion protein bound to IgG-sepharose resin was set aside for ternary complex assembly. (Further details are provided in the Materials and Methods section 7.5.) Amgen (Thousand Oaks, USA) kindly supplied the recombinant mature human GDNF, which they produced in *Escherichia coli*.

### 3.2.3 Assembly of the mTC

The rGFRα1 was produced with a C-terminal TEV cleavable Protein A tag. This fusion protein immobilised on IgG-sepharose and formed the basis for mTC assembly. Briefly, to assemble the mTC complex excess hGDNF (Amgen, Thousand Oaks) was added to the immobilised rGFRα1, allowed to bind for one hour, and then unbound protein was removed by washing. Affinity-purified hRET-ECD was subsequently added to the immobilised rGFRα1/hGDNF complex in the presence of 1 mM calcium chloride and allowed to bind for one hour at 4 °C. The assembled IgG-sepharose-bound mTC was then washed with mTC buffer to remove any unbound hRET-ECD. Finally the mTC was eluted from the IgG-sepharose by addition of TEV protease to cleave off the rGFRα1 Protein A tag and liberate the mTC from the resin. The mTC was subsequently purified by size exclusion chromatography and the complex stoichiometry assessed by SDS PAGE and TapeStation (Agilent) analysis. This confirmed equimolar of the three component proteins were present in the purified complex (Figure 3.1a). The theoretical molecular weight of the 2 hRET-ECD/2 rGFRα1/1 hGDNF dimer mTC is 264.4 kDa plus sugar.

### 3.2.4 Negatively stained EM grid production and image collection

mTC samples from individual 1 ml size exclusion fractions were applied to carbon grids and negatively stained with 2% (w/v) uranyl acetate. The 'best' fraction was selected by visual inspection on the basis of sample homogeneity. Images of the selected grid for subsequent analysis were taken on a Tecnai T20 electron microscope, equipped with a 4000×4000 pixel CCD detector, at 80000× magnification giving a sampling of 2.17 angstroms/pixel (Figure 3.1a).

### 3.2.5 Particle picking and class generation

Particle images were manually picked from EM grid images with a 192×192 pixel box size (416.64 $Å^2$) using Boxer, part of the EMAN suite (Ludtke *et al.*, 1999). A stack of 11000 particles was generated. The particle images were coarsened to 4.34 Å/pixel, high pass filtered to 200 Å in IMAGIC (van Heel *et al.*, 1996), normalised and initially low pass filtered to 30 Å in SPIDER (Frank *et al.*, 1996). Given the low signal to noise

ratio in EM images of proteins, particles images of the same view are aligned, assigned to classes, and averaged to increase the signal to noise ratio. Classes were initially generated using EMAN reference-free alignment and classification (Ludtke *et al.*, 1999), and subsequently by alignment to the 3D reconstruction using SPIDER (Frank *et al.*, 1996) and classification in IMAGIC (van Heel *et al.*, 1996). Two distinctive class averages were immediately apparent: a 'figure of eight' view (third class in Figure 3.1c), containing an apparent two-fold axis, and a 'leg' view (sixth class in Figure 3.1c).

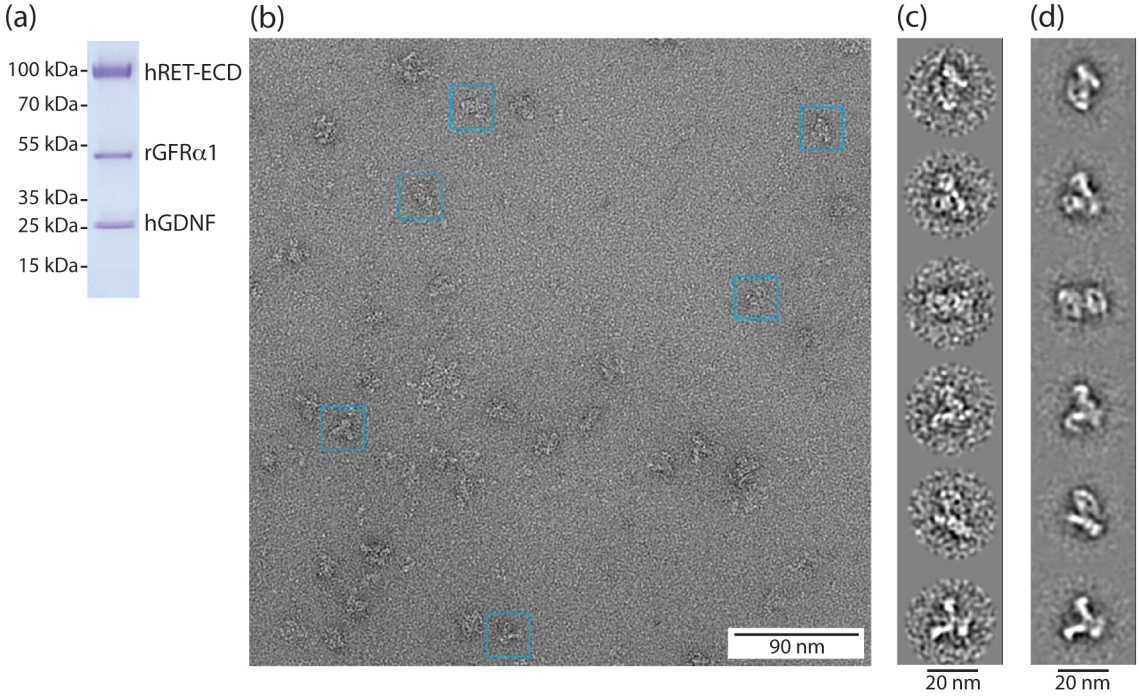



**Figure 3.1 Negative stain EM of the mTC**

(a) SDS-PAGE of the mTC following size exclusion chromatography with a Superose 6 column (GE Healthcare). (b) Representative image of the mTC EM grid negatively stained with 2% (w/v) uranyl acetate. (c) Representative picked particle images after filtering and normalisation. (d) Corresponding representative class averages.

### 3.2.6 Producing a 3D reconstruction by angular reconstitution

Multi-variate statistical analysis (MSA) (van Heel, 1984) of the centred particle image stack was carried out to generate eigenimages, which indicated the particle likely possessed two-fold symmetry, commensurate with the apparent two-fold in the 'figure-of-eight' view (Figure 3.2a). The eigenimages also indicated possible three-fold symmetry. However further analysis revealed these eigenimages resulted predominantly from the randomly rotationally orientated 'leg' view particles, which are not three-fold

symmetric, but since the legs are at a 120° angle they give rise to an apparent three-fold in the eigenimages when these particles are randomly rotationally orientated. The 3D reconstruction was therefore produced applying C2 symmetry, using the angular reconstitution method (van Heel, 1987) in IMAGIC.

The 3D reconstruction process was started using a two-fold symmetrical 'figure of eight' view and a 'leg' view shown in Figure 3.2b, since these seemed likely to be orthogonal views. In the 'start-up' of the reconstruction using IMAGIC these views were assigned to be around 100° apart. These two class averages were then used as anchors (or references) to assign Euler angles to a larger selection of class averages. Generation of a 3D reconstruction based on this Euler assignment, and visual comparison of re-projections from this 3D with the input class averages allowed a small selection of classes with apparently correctly assigned Euler angles to be defined. These were then used as anchors for reassignment of Euler angles of the larger selection of class averages. Once again, correctly assigned class averages were selected and used as anchors for the next round of Euler assignment and 3D reconstruction generation. This process was then repeated until no new class averages were added to the anchor set. The resulting 3D reconstruction from the correctly assigned class averages was then used as a reference for the generation of new classes from the particle stack. These new class averages were assigned Euler angles using the anchor set, and the iterative process of trying to increase the number and range of class averages used in the 3D reconstruction repeated.

Initial attempts using this method were unsuccessful at getting beyond a very small selection of correctly Euler angle assigned class averages. Alternative pairs of classes for the initial start-up were also tried, but this was also unsuccessful.

Critical assessment of the stack of particle images showed a wide distribution of stain levels, and many poorly defined particle images. Since this variability and poor signal to noise ratio will negatively impact particle image alignment and class assignment, a stack of around 2000 well stained clearly defined particles images was selected. It was hoped that this would result in the production of better classes, due to improved particle alignment and class assignment, which would then aid the production of a 3D reconstruction.

Classes were generated from this new stack by alignment to forward projections of the best previous 3D reconstruction, followed by MSA based class assignment. Repeating the start-up, and anchor set development procedure using these class averages was successful. Fresh classes were then generated by alignment of the particle images to forward projections of the new 3D reconstruction followed by MSA class assignment. The resulting class averages were then assigned Euler angles using the previous anchor set, and a new 3D reconstruction was produced from the correctly assigned class averages. This process of improved class generation by alignment of the stack to the most recent 3D, alternated with Euler assignment and generation of a new 3D reconstruction, was iteratively repeated until no further improvement in the class averages or the 3D was observed.

### 3.2.7 Refining the 3D reconstruction using projection matching

The angular reconstitution generated 3D reconstruction was subsequently refined by projection matching. For this step the entire 11000 particle image stack was used, since fresh sample was not available (due to expression problems with rGFRα1) to expand the more critically selected stack. For this step the low-pass filter was changed to 18 Å. The forward projections from the angular reconstitution generated 3D reconstruction were much better references for particle alignment and class assignment, than those used in the initial development of the 3D reconstruction. Therefore, reasonable particle alignment and class assignment even with the less well-defined particle images was observed, allowing the use of this larger particle image stack.

The projection matching procedure involved Euler angle assignment of classes based on matching to forward projections of the current 3D reconstruction, which was carried out using IMAGIC. Correctly Euler angle assigned class averages were selected by manual comparison between the class and the matched forward projection. These correctly Euler angle assigned classes were then used to generate a 3D reconstruction, which was then used as the reference for the next round of class generation and projection matching. This procedure was repeated until the range of class averages included in the reconstruction, and the quality of the match between class averages and re-projections ceased to improve. Details of the final 3D reconstruction at 27.8 Å resolution are shown in Figure 3.2, and images of the 3D reconstruction are shown in Figure 3.3.

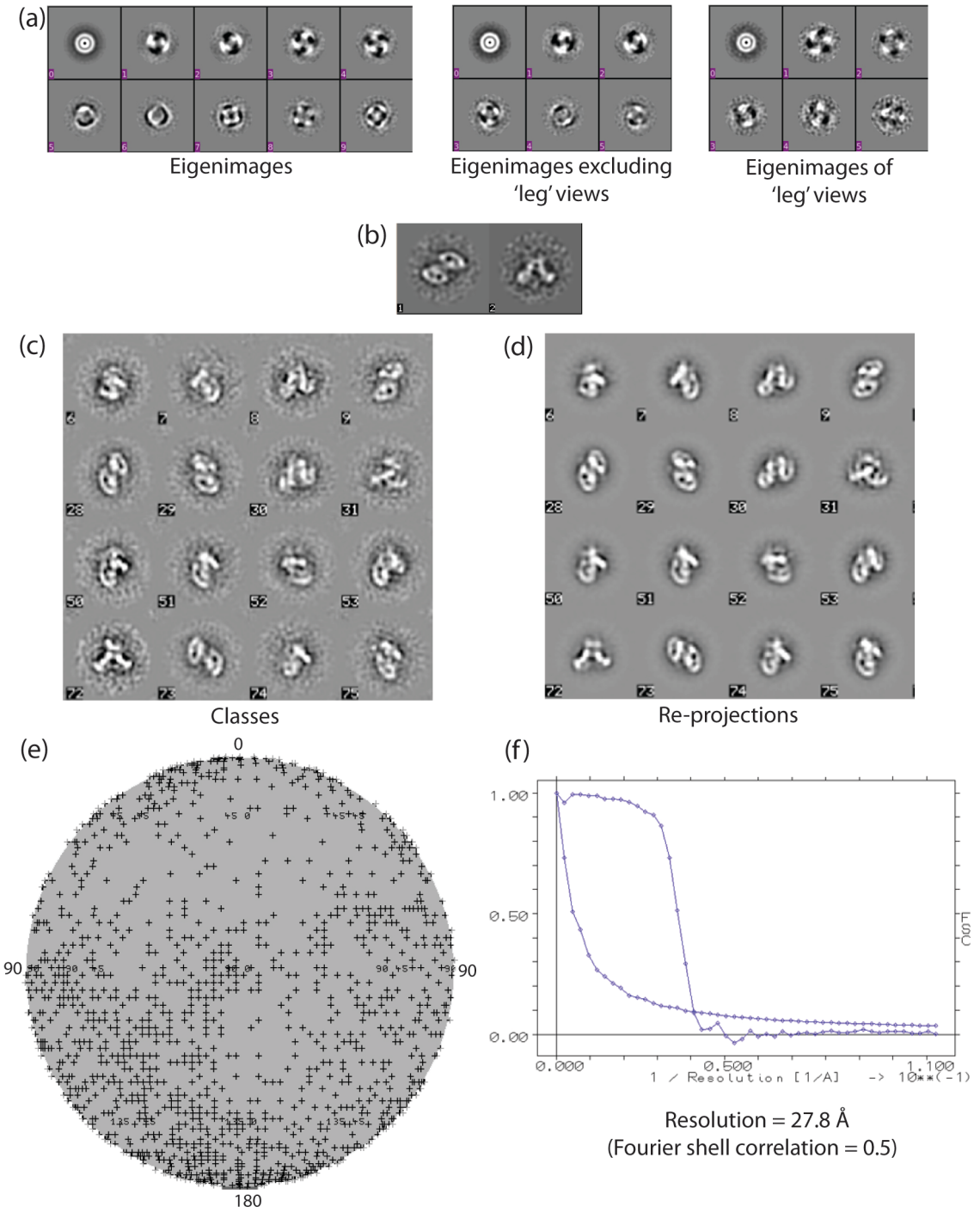



**Figure 3.2 Details of the mTC 3D reconstruction**

3D reconstruction of the mTC from the 11000 particle stack. (a) Eigenimages of the whole stack and two subsets as annotated. (b) The two near orthogonal class averages used to initiate the 3D reconstruction process are shown. (c) A selection of input class averages (classes) and (d) the corresponding re-projections from the final 3D reconstruction are shown. (e) Map of Euler angles assigned to the input class averages in the final 3D reconstruction. (f) Fourier shell correlation to determine map resolution.

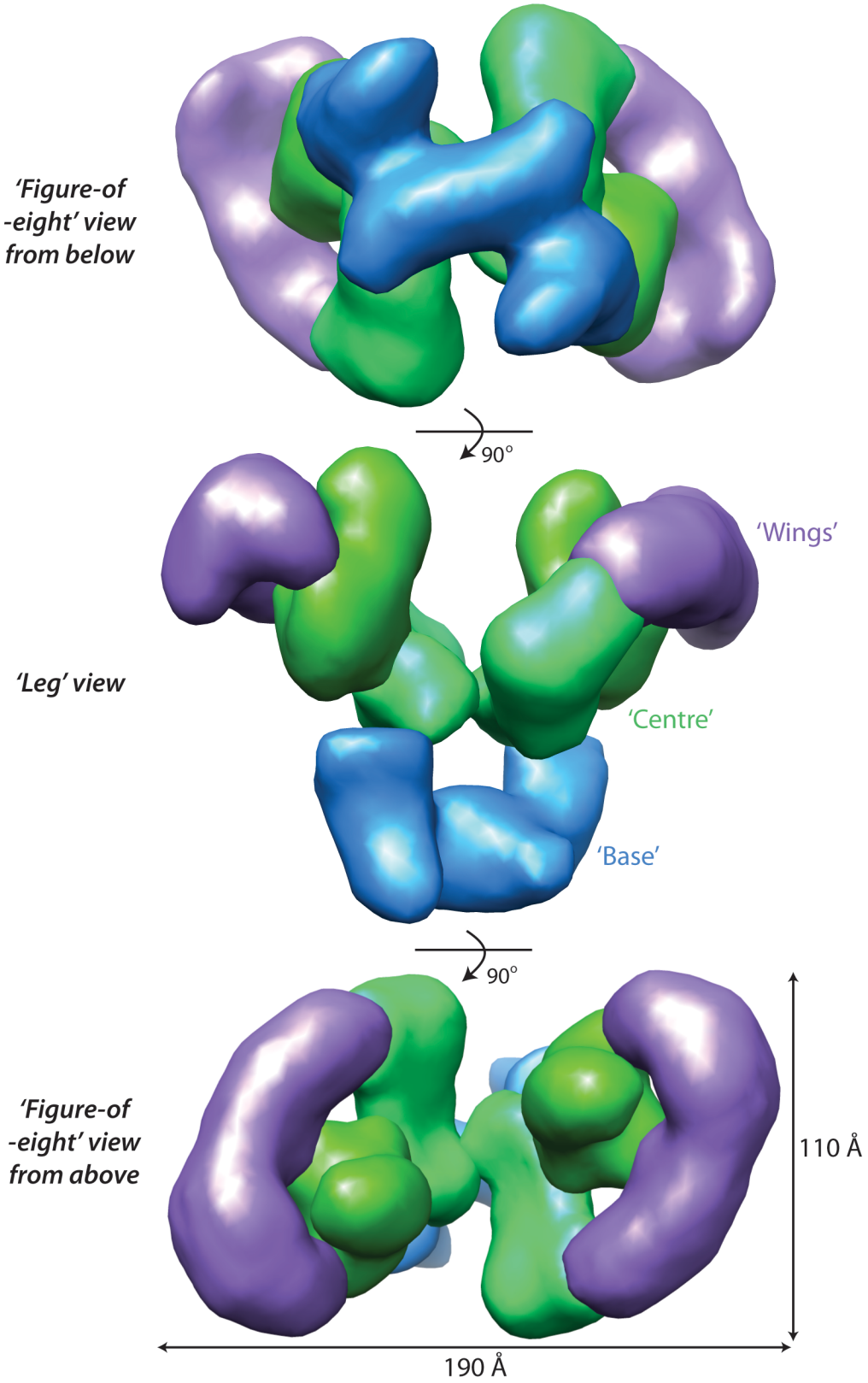



**Figure 3.3 Mammalian ternary complex 3D reconstruction**

Surface view of the mTC reconstruction, with the 'wing', 'centre' and 'base' regions (discussed in the text) coloured purple, green and blue respectively.

## 3.3 Producing a model of the mammalian ternary complex based on the EM 3D reconstruction

### 3.3.1 Known structures and models of components

Crystal structures of hRET-CLD(1–2) (Kjaer *et al.*, 2010) and the hGDNF/rGFRα1-D2D3 (Parkash *et al.*, 2008) complex have been published previously. As described in Chapter 2, I have produced a SAXS-based model for the zRET-ECD defining the arrangement of the four CLDs. GFRα1 D1 is reliably predicted to have the same fold as the GFRα D2 and D3 (Leppanen *et al.*, 2004), and can therefore be modelled with reasonable confidence using the PHYRE2 structure prediction program (Kelley and Sternberg, 2009). This leaves the RET-CRD, the 44 residue linker between GFRα1 D1 and D2, the 75 residue C-tail of GFRα1 and the 39 residue N-terminal tail of GDNF, for which there is no structural information and which cannot be reliably modelled. As described in section 2.10.4, the 126 residue RET-CRD lacks regular secondary structure elements but is folded and likely defined by the 8 (in hRET) predicted disulphide bridges. The other un-modelled elements are all predicted to be unstructured tails/linkers (Figure 3.4). If any of these elements are completely randomly arranged they will not contribute to the EM 3D reconstruction, since they would have been 'averaged out' in the class averages.

### 3.3.2 Assumptions used in fitting mTC components into the EM map

These crystal structures, homology models and the SAXS-derived RET-ECD model were docked into the final 3D reconstruction map (Figure 3.3) using CHIMERA in order to generate a preliminary model of the arrangement of the mTC components. This fitting was carried out based on the following assumptions:

First, the intrinsic two-fold symmetry of the GDNF dimer was assumed to be co-incident with the mTC two-fold symmetry. The hGDNF/rGFRα1-D2D3 complex crystal structure also exhibits two-fold symmetry and a 2:2 stoichiometry. Therefore it seemed reasonable to assume the ligand/co-receptor two-fold axis is coincident with the mTC EM map two-fold axis.

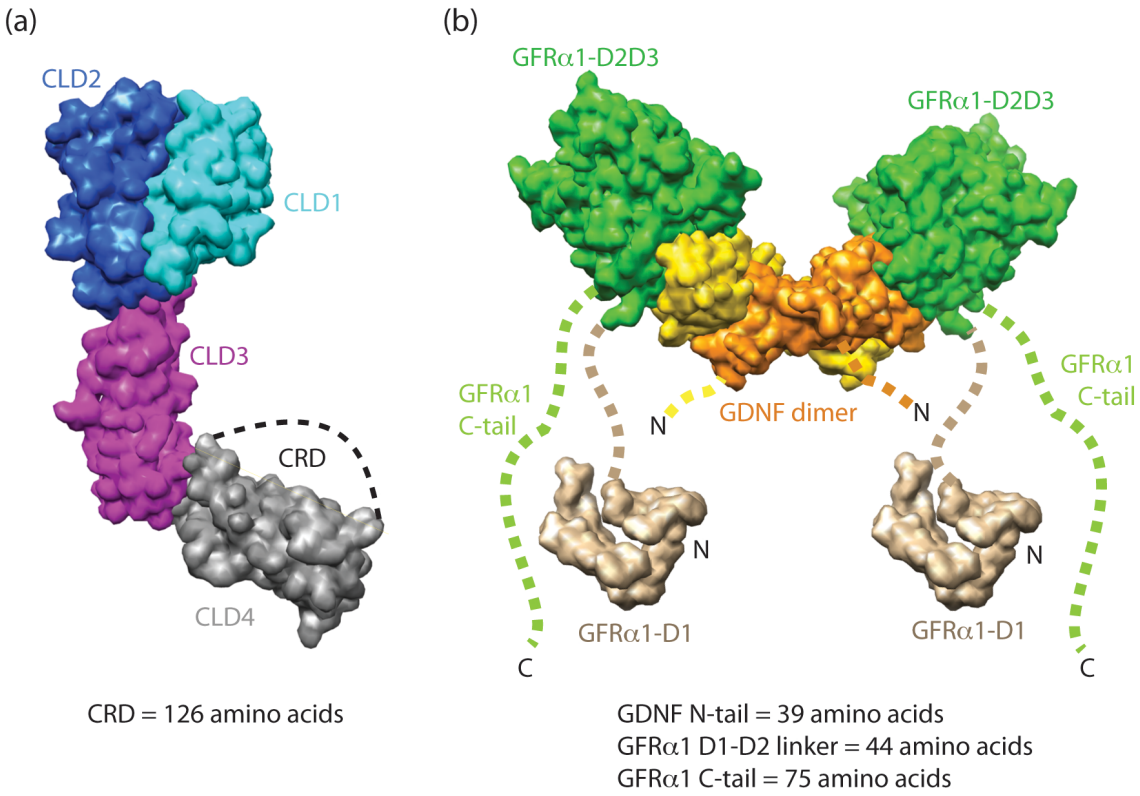



**Figure 3.4 Structural models for the RET ternary complex components**

(a) SAXS-derived model of zRET-ECD, with CLDs shown in surface view at 4 Å. It is assumed that the overall arrangement of the CLDs in hRET is equivalent to zRET. (b) Surface view at 4 Å of the hGDNF/rGFRα1-D2D3 crystal structure (PDB 2V5E; Parkash *et al.*, 2008) and structure prediction model of hGFRα1-D1. Predicted unstructured elements of GFRα1 and GDNF are shown schematically as dashed lines.

Second, the arrangement of the human RET CLDs is assumed to a lesser or greater extent to match that seen in the zebrafish RET-ECD SAXS-based model. The conservation of the structural elements that define the arrangement of the CLDs in RET between human and zebrafish was discussed in the previous chapter (section 2.2). The basis for assuming that the arrangement of the CLDs is mostly maintained between the RET-ECD alone, as defined by SAXS, and the RET-ECD when bound to ligand/co-receptor in the mTC are as follows: The clamshell arrangement of CLD(1–2), as seen in the hRET-CLD(1–2) crystal structure, is assumed to be maintained in the mTC since the arrangement is tethered by a conserved disulphide between CLD1 and CLD2. The CLD2–CLD3 arrangement is maintained by calcium ions coordinated by conserved calcium binding residues at the CLD2–CLD3 junction. Calcium is known to be required for ternary complex formation (Anders *et al.*, 2001; Nozaki *et al.*, 1998), thus we

assume the CLD(1–3) interdomain angles will be conserved and maintained. However the 100° bend angle between the CLD3 and CLD4, as determined by my SAXS analysis, cannot be assumed to be maintained in the mTC with such confidence. The linker between CLD3 and CLD4 is conserved between human and zebrafish RET, but there is no reason to think it may not be subject to a conformational change upon ligand/co-receptor binding. Furthermore, the possible apparent rotation of CLD4 between the SAXS-based models of zRET-CLD(1–4) and zRET-ECD, creates another layer of uncertainty about the arrangement of this interdomain junction in the ternary complex. As a starting point the zRET-ECD SAXS-based model including the CLD3–CLD4 arrangement was used without alteration to fit into the 3D reconstruction, bearing in mind the possibility of variation in the CLD3–CLD4 arrangement.

Finally, the arrangement of the hGDNF/rGFRα1-D2D3 complex was assumed to be the same as that observed in the crystal structures. Comparison between the published crystal structures of hGDNF showed that the intermonomer angle in the disulphide-linked dimer can vary from 146–168° (Parkash and Goldman, 2009). However, despite differences in the hGDNF structure in the two published hGDNF/rGFRα1-D2D3 complexes (PDB 2V5E (Parkash *et al.*, 2008) & 3FUB (Parkash and Goldman, 2009)) the overall organisation of the complex is very similar with the distance between the putative RET binding sites on GFRα1 near identical (Parkash and Goldman, 2009). Both of these crystal structures were used in fitting and 2V5E, which has a GDNF intermonomer bend angle of 160°, was chosen based on a slightly improved fit for GDNF. The assumption that the arrangement of the hGDNF/rGFRα1-D2D3 complex is unchanged from this crystal structure in the mTC is reasonable for fitting in the first instance given the low resolution of the EM map, but the known flexibility of GDNF means that the precise arrangement may deviate from the PDB 2V5E structure by up to 20°.

### 3.3.3 Fitting the mTC components into the EM map

Based on these assumptions and considerations, the models depicted in Figure 3.4 were fitted into the 3D reconstruction map. Fitting was carried out manually in CHIMERA followed by use of CHIMERA's 'fit-in-map' tool whenever possible. To arrive at the

most appropriate interpretation of the arrangement of the RET ternary complex components we considered a number of different interpretations. The best fit for the GDNF/GFRα1-D2D3 crystal structure was found to be in the centre of the map co-incident with the 2-fold (correlation coefficient = 0.81, Figure 3.5). This left the possibility of the RET-ECD either being on the 'inside' of the 3D reconstruction, occupying predominantly the base and central regions of the map or located wrapped around the 'outside' of the object, in the wing regions of the map (the different regions are annotated in Figure 3.3). The best interpretation, based on the correlation coefficient between the map and the fitted RET-CLD(1-4), was found to be with RET-ECD in the wings of the complex, which I will discuss below. Further on in this chapter I will also discuss one of the considered RET 'inside' interpretations, since it has important ramifications for the lack of involvement of the previously observed crystallographic CLD(1–2) dimer in the mTC.

### 3.3.4 Interpretations of the 3D reconstruction with external RET-ECD

With the GDNF/GFRα1-D2D3 crystal structure fitted in the centre of the map, the 'wings' of the 3D reconstruction are the largest stretches of unexplained well-connected density. It is striking how well the zRET-ECD SAXS-model fits into these 'wings' without any alterations to the CLD arrangement. It is not a unique solution however, since either orientation of the RET-ECD model can be accommodated within the 3D map with the same correlation coefficient (0.82). I define each orientation as 'central CLD4' (Figure 3.5a) and 'central CLD(1–2)' (Figure 3.5b). Positioning RET with the CLD4's towards the centre of the map, as in the 'central CLD4' model, is also the most logical of the two possible RET configurations since there is available density near to CLD4 for the CRD to occupy. This would position the CRDs near the base of the complex close to one another, consistent with the transmembrane domains interacting in the native full-length complex (Kjaer *et al.*, 2006). This central CLD4 model was therefore selected as the current best interpretation since all the components with known structure fit well within the map, and the remaining unexplained density could readily accommodate the remaining elements (RET-CRD; GFRα-D1; GFRα-Ctail). The implications of this 'central CLD4' interpretation are discussed below.

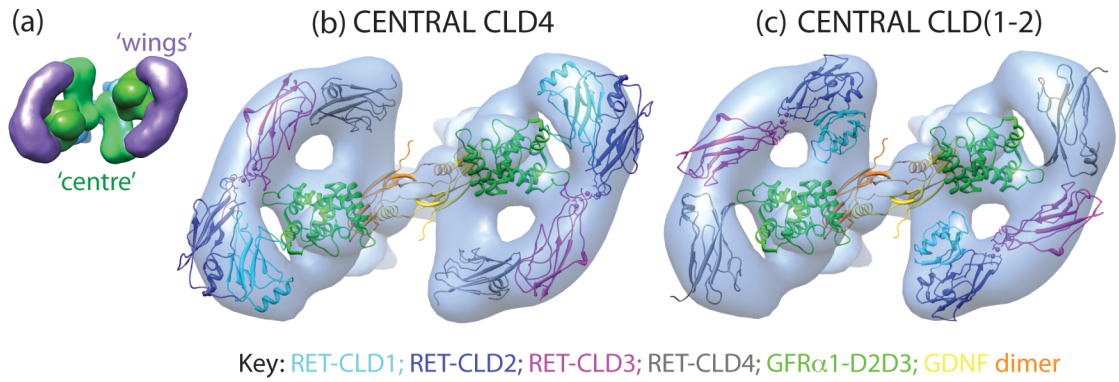



**Figure 3.5 Fit of the RET-ECD into the wings of the 3D reconstruction**

(a) Surface view of the mTC reconstruction, with the 'wing', 'centre' and 'base' regions coloured purple, green and blue respectively. (b)&(c) Transparent surface view of the mTC 3D reconstruction, with fitted RET, GFRα1 and GDNF structures shown in ribbon representation. Two alternative orientations of RET are shown in (b) and (c).

### 3.3.5 Implications of the 'central CLD4' interpretation of the mTC 3D reconstruction

Of the possible models based on the assumptions and considerations described in section 3.3.2, the best model places the ligand/co-receptor complex in the centre of the complex, with RET-CLD4CRD drawn together towards the base of the complex, and CLD(1–3) splayed apart wrapping around the ligand/co-receptor (Figure 3.6). This arrangement results in multiple contacts between RET and both GDNF and GFRα1-D2D3, and also predicts the location of the unmodelled elements.

#### *3.3.5.1 RET-CLD(1–3):GFRα1-D3 contacts*

Intriguingly, the mammal-specific CLD1 surface, involved in the hRET-CLD(1–2) crystallographic dimer, is contacting the GFRα1-D3. The CLD(1–2)-CLD3 junction is also weakly contacting the co-receptor. This may explain both the calcium dependence of the interaction with ligand/co-receptor, and the higher vertebrate species specificity conferred by CLD1, 2 and 3 in the ligand/co-receptor interaction (Kjaer and Ibanez, 2003a).

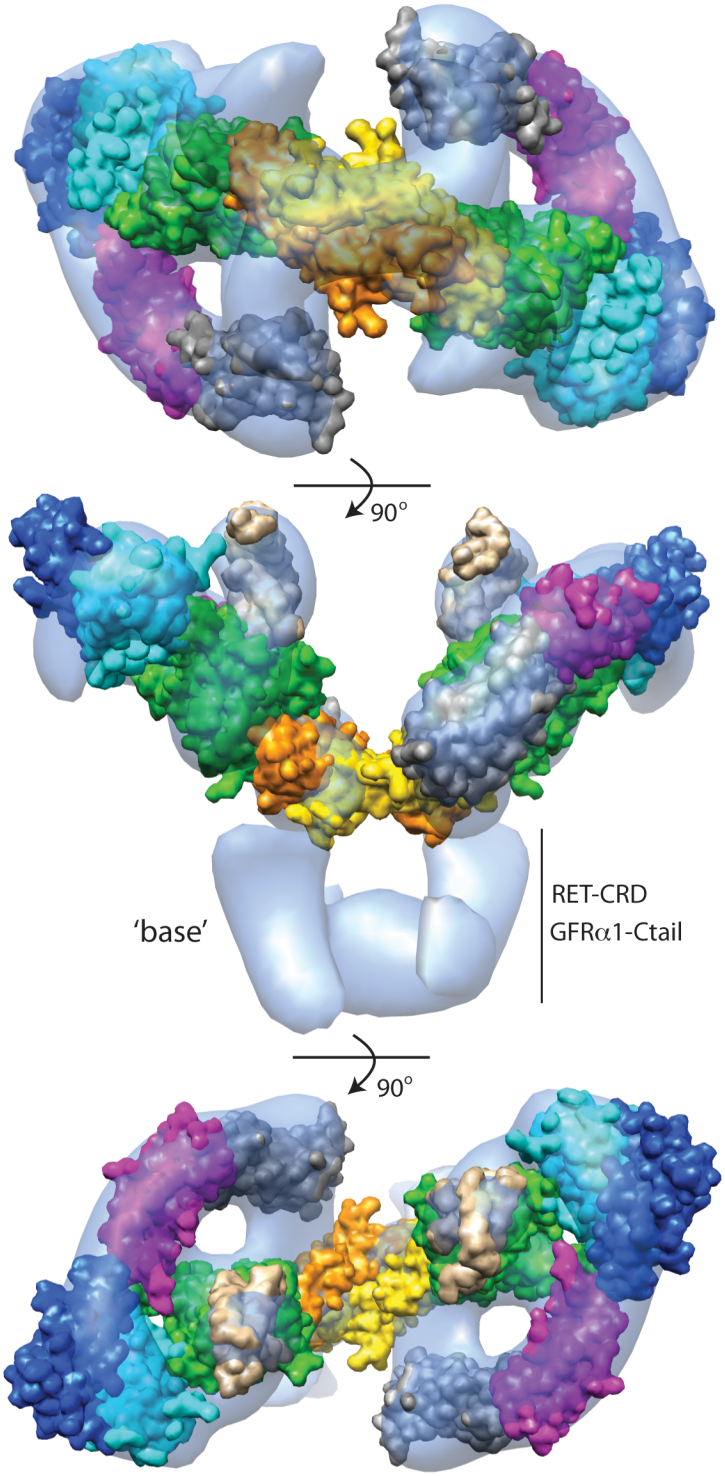



**Figure 3.6 Best interpretation of the EM map**

Transparent surface view of the mTC 3D reconstruction, with fitted RET, GFRα1 and GDNF structures shown in surface representation at 4 Å.

#### 3.3.5.2 RET-CRD:GDNF contacts

CLD4 is close to the ligand/co-receptor junction, and by extension the CRD is presumably, at least in part, occupying the available density adjacent to CLD4 engaging GDNF and GFRα1-D2. This places the CRD, at least partly, in a similar position to that observed by SAXS folded back against CLD4. The central location of the CRD may explain unpublished data by Dr. Svend Kjær, showing that two antibody epitopes within the CRD become inaccessible upon ternary complex formation. This is also in agreement with crosslinking data, which suggested the CLD4CRD junction and GDNF N-terminal tail are in close proximity in the mTC.

#### 3.3.5.3 The position of GFRα1-D1

The GFRα1-D1 model fits neatly into the density above the GFRα1-D2D3's observed in the leg view as shown in Figure 3.6. It could also fit in the base region of the 'leg' view adjacent to the CRD in this interpretation. Since it is linked to D2 by a fairly long linker, there aren't any known steric constraints on its position. I have shown it above GFRα1-D2D3 as the best explanation for that region of density. However, since D1 has been shown to crosslink with the CRD and confers stability to the mTC, it may be more consistent with these data to place it in the base region engaging the CRD.

#### 3.3.5.4 The membrane facing elements: RET-CRD and GFRα1-Ctail

In this interpretation the RET-CRDs are presumed to be engaging GDNF in the central region of the map, as well as extending into the base region towards the putative position of the cell membrane. Since the structure of the CRD is unknown this is purely speculative. However the juxtaposition of the CRDs in this interpretation is consistent with the dimerisation of the TMs in the full-length RET context.

I have placed the 75 residue C-tail of GFRα1 in the unexplained base region density, since it would face the membrane along with the CRD in the membrane-localised full-length complex. However given that this tail has no predicted secondary structure, and contains a long string of threonine residues and only a single potential intramolecular disulphide, it may be too flexible to contribute much to the 3D reconstruction.

#### *3.3.5.5 The hand of the EM map has not been determined*

A tilt pair analysis to determine the hand of the EM map *de novo* (Rosenthal and Henderson, 2003) has not been conducted as yet. Fitting was therefore carried out using both the map shown here and its enantiomorph (generated in IMAGIC). We had hoped that the fitting would unambiguously reveal which was the correct hand. Unfortunately, at the current resolution of the map the fit of the components is comparable between the map and its enantiomorph. The correlation coefficient of RET-CLD(1-4) is 0.78 in the enantiomorph map and 0.82 in the displayed map and the correlation coefficient of GDNF/GFRα1-D2D3 is 0.80 in the enantiomorph vs 0.81. The central CLD4 arrangement of the components is the same in the map and its enantiomorph, although the precise interaction sites between components differ. RET-CLD1 and the CLD(1–2)-CLD3 junction still contact GFRα1-D3 in the enantiomorph model. However the interfaces are swapped with the region on GFRα1-D3 that contacts CLD1 in the model described above contacting the CLD(1–2)-CLD3 junction in the enantiomorph model, and vice versa. Similarly the RET-CRD still contacts GDNF, but on an opposing surface of GDNF. The map used in the figures above was chosen for two reasons. Firstly, the previously proposed RET-binding surface on GFRα1 faces inside the complex in the fit into this map, whereas it is facing outside the complex in the enantiomorph map fit. Secondly, in the enantiomorph map the central CLD4 arrangement results in the RET-CRD contacting the longest helix in GDNF (the GDNF 'heel'). This helix contains a centrally located conserved N-linked glycosylation site, making it unlikely to be a protein interface site. These reasons are not conclusive however, and improved map resolution to unambiguously determine the hand by fitting or tilt-pair analysis to determine the hand *de novo* will be needed before the contact sites between components in the ternary complex can be discussed with more confidence.

### 3.3.6 The CLD(1–2) crystallographic dimer does not fit in the EM map

The crystal structure of hRET-CLD(1–2) revealed a crystallographic dimer. Mutation of the dimer interface reduced the apparent affinity of immobilised hRET-ECD for ligand/co-receptor ten-fold, and therefore formation of the hRET-CLD(1–2) dimer was

deemed to be important for ternary complex stability (Kjaer *et al.*, 2010). However the RET-ECD fit into the EM map was poor (correlation coefficient = 0.53) if the dimer arrangement was imposed.

The CLD(1–2) dimer, like the ligand/co-receptor complex, possesses a 2-fold axis, which was therefore assumed to be coincident with the EM map two-fold in fitting. The presence of the CLD(1–2) mediated dimer in the context of the CLD(1–4) model, results in a 'tweezer' shaped object (Figure 3.7a). The fit of this object in the EM map is shown in Figure 3.7b (correlation coefficient = 0.53). It is a poor fit however since CLD3 is largely outside of the map. Furthermore it leaves a large amount of unexplained density, possibly far too much to be explained by the unmodelled elements or by a different CLD3–4 arrangement.

The lack of a good fit of the CLD(1–3) dimer into the EM map strongly suggests that the crystallographic dimer is not present in the ternary complex. We therefore sought an alternative explanation for the impact of the R77E/R144D dimer interface mutations on ternary complex assembly, and the role of the CLD1 elements in conferring ligand/co-receptor specificity (discussed in section 1.6.4.2). One possibility is that, as originally proposed by Kjær *et al.* (Kjaer and Ibanez, 2003a), these elements are directly involved in binding ligand/co-receptor, as indeed is observed in the current interpretation.

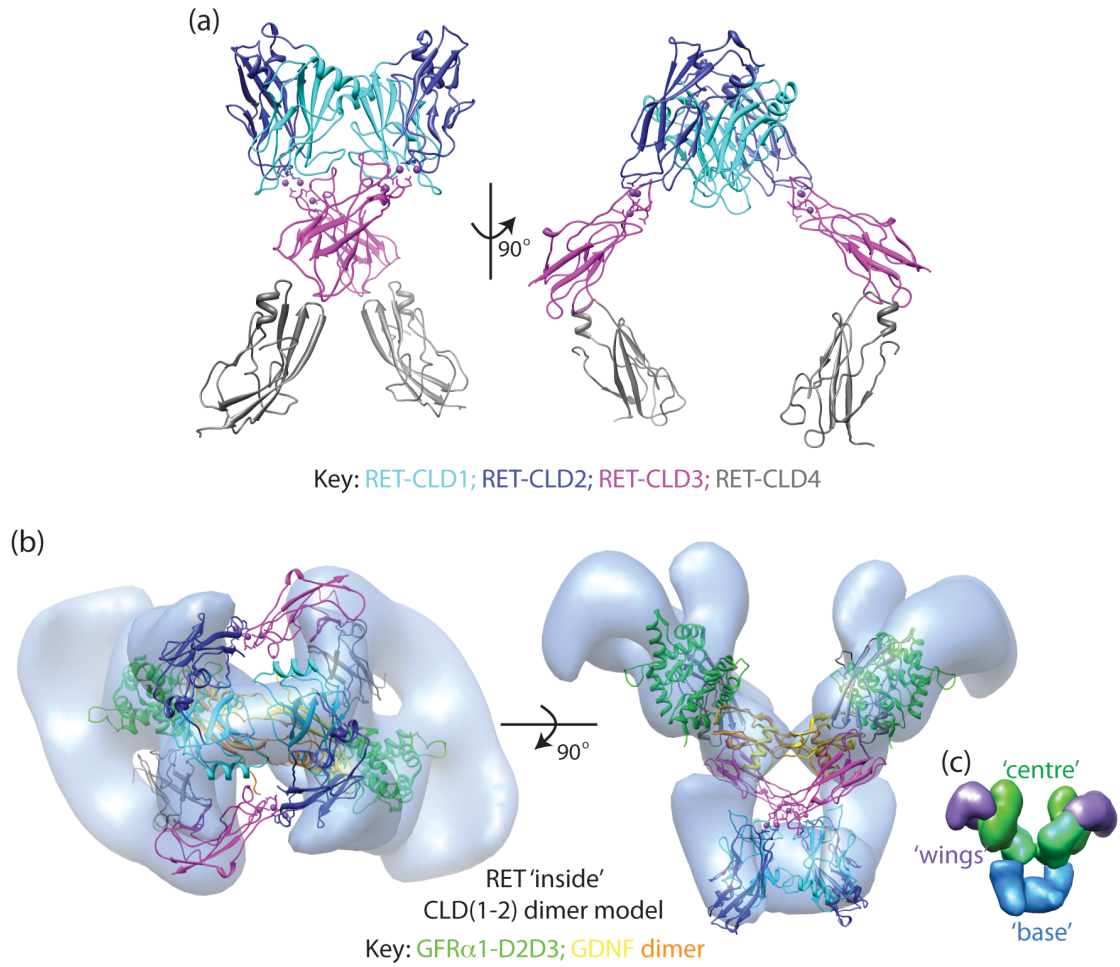



**Figure 3.7 Imposing the CLD(1–2) crystallographic dimer on the SAXS based zRET-CLD(1–4) arrangement and fitting in the mTC EM map**

(a) Ribbon representation of the SAXS-derived zRET-CLD(1–4) model with the CLD(1–2) dimer, observed in the hRET-CLD(1–2) crystal structure (Kjaer *et al.*, 2010), imposed. (b) Transparent surface view of the mTC 3D reconstruction, with fitted RET, GFRα1 and GDNF structures shown in ribbon representation. The best fit of the components with the CLD(1–2) crystallographic dimer imposed is shown. (c) Surface view of the mTC reconstruction, with the 'wing', 'centre' and 'base' regions coloured purple, green and blue respectively.

## 3.4 Aids in interpreting the EM map

There are a number of possible experiments that can be conducted to validate interpretations and help identify the location of components within the EM map of macromolecular complexes. These include domain deletions, addition of an EM-visible protein domain or tag to particular components or improvement of the resolution of the 3D map. Our current strategy is to try to improve the resolution of the EM map both by

increasing the number of particle images used in the negative stain reconstruction and hopefully moving to cryo-EM. Dr. Svend Kjær has recently produced a better quality sample of the mTC using an additional ion exchange chromatography step for purification. This additional purification step has improved sample homogeneity, as judged by visual analysis of negatively stained EM grids, and this sample is currently being analysed and worked up into a new 3D reconstruction.

### 3.4.1 Assessing the role of the GFRα1-D1 in the mammalian ternary complex

The rGFRα1-D1 is only 88 residues (10 kDa) in size and therefore is not a particularly good marker within a negative stain EM map. However its location is ambiguous and it is the only structured domain believed to be dispensible for ternary complex formation, although it may enhance mTC complex stability (Virtanen *et al.*, 2005). A rGFRα1$_{\Delta D1}$ construct was produced, corresponding to residues 150–427 to test this hypothesis. Initially it appeared that the ternary complex could still be assembled using this truncated co-receptor immobilised on IgG-sepharose resin. However size exclusion chromatography of the complex showed peaks only for the individual component proteins, and not the intact mTC complex. This was confirmed by inspection of negatively stained EM grids prepared from fractions from the highest MW peak, which showed either aggregates or very small particles (Figure 3.8). This complex is therefore too unstable for EM analysis, indicating rGFRα1$_{\Delta D1}$ may be indeed be important for the stability of recombinant mTC.

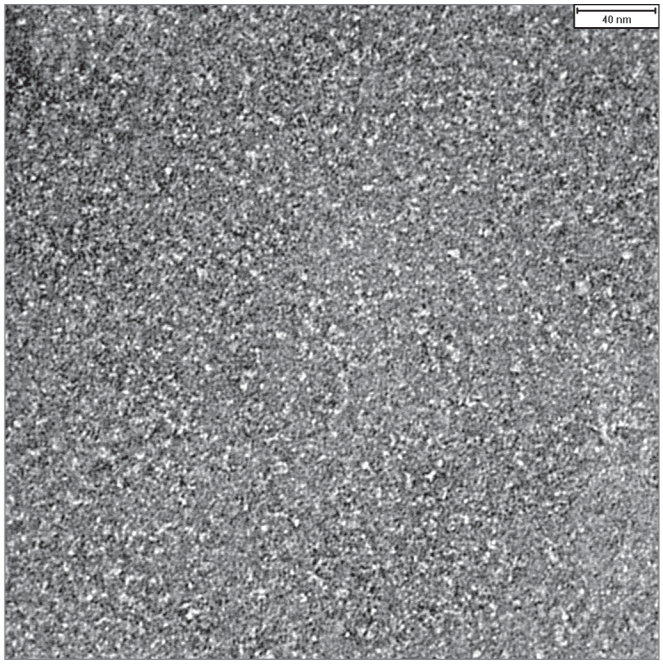



**Figure 3.8 Negatively stained EM grid of GFRα1$_{\Delta D1}$ mTC sample**

EM grid negatively stained with 2% (w/v) uranyl acetate. No particles of a similar size to the intact mTC are evident.

## 3.5 Conclusions

In summary, we have produced a 27.8 Å resolution negative stain EM 3D reconstruction of the mTC, which has internal two-fold symmetry. Preliminary fitting of known structures and homology models into this EM map resulted in an interpretation for the arrangement of the RET-ECD, GDNF and GFRα1 in the mTC. This interpretation shows the GDNF dimer/ 2 GFRα1 complex lies at the heart of the mTC, with the RET-ECD wrapped around the outside. The RET-CRDs are drawn together at the base of the complex and also make contacts with GDNF. The RET-CLD(1–3) regions are splayed apart from each other interacting with GFRα-D3 at the distal ends of the GDNF/GFRα1 complex. Surprisingly, we do not observe the crystallographic RET-CLD(1–2) dimer within the mTC 3D reconstruction. Instead, CLD1–2 interface, previously shown to be important for ternary complex stability, appears to be involved in contacting GFRα1-D3 rather than in mediating a CLD(1–2) dimer in the mTC.

# Chapter 4. Towards structural analysis of a lower vertebrate RET's interaction with ligand/co-receptor

## 4.1 Chapter aims

As detailed in the introduction, human RET-CLD(1–2) has mammal-specific structural elements, which are required for interaction with mammalian ligand/co-receptor complexes (section 1.6.5). Homologue-scanning mutagenesis studies, combined with the hRET-CLD(1–2) crystal structure, revealed an unexpected diversity separating higher and lower vertebrates both at the sequence and structural levels (Kjaer *et al.*, 2010; Kjaer and Ibanez, 2003a). In this thesis I have sought to define and understand these differences by looking at a lower vertebrate (zebrafish) RET alongside the human orthologue. Towards this aim I endeavoured to produce a zebrafish ligand and co-receptor in order to characterise their interaction with zebrafish RET by structural methods as a comparison with the mammalian complex described in the preceding chapter.

A second rationale was that the amount of mammalian RET ternary complex we were able to produce was very limited even from large-scale mammalian cell expression. We were therefore not able to pursue crystallisation of this complex. An additional motivation for developing a protocol for production of a zebrafish ligand and co-receptor was therefore to try to obtain sufficient quantities of each of the zRET ternary complex components to initiate crystallisation trials of the zRET ternary complex. Since atomic resolution of protein structures can be obtained more readily by X-ray crystallography compared to EM (excepting cryo-EM analysis of very high quality and/or high symmetry samples) it was hoped this approach would provide a much more detailed description of the interactions between RET and ligand/co-receptor.

## 4.2 Vertebrate ligand and co-receptor sequences and alignments

Alignment of the amino acid sequences of the GFRα1 co-receptor reveals high sequence identity between lower and higher vertebrate GFRα1, except between the membrane proximal C-tails. Overall the sequence identity between human GFRα1 and zebrafish GFRα1a is 63%. The identity without the C-tail is 73%, with all three structured domains (D1–D3) and the D1–D2 linker showing high conservation (Figure 4.1).

Alignment of the GDNF amino acid sequences from lower and higher vertebrate species shows that the sequence of the mature protein, excluding the pro-sequence, is highly conserved, particularly in the structured region (Figure 4.2). The sequence identity of the mature protein between human and zebrafish is 60%, rising to 69% in the structured region. The N-terminal tail of zebrafish GDNF contains a fish specific highly charged 14 amino acid insertion.

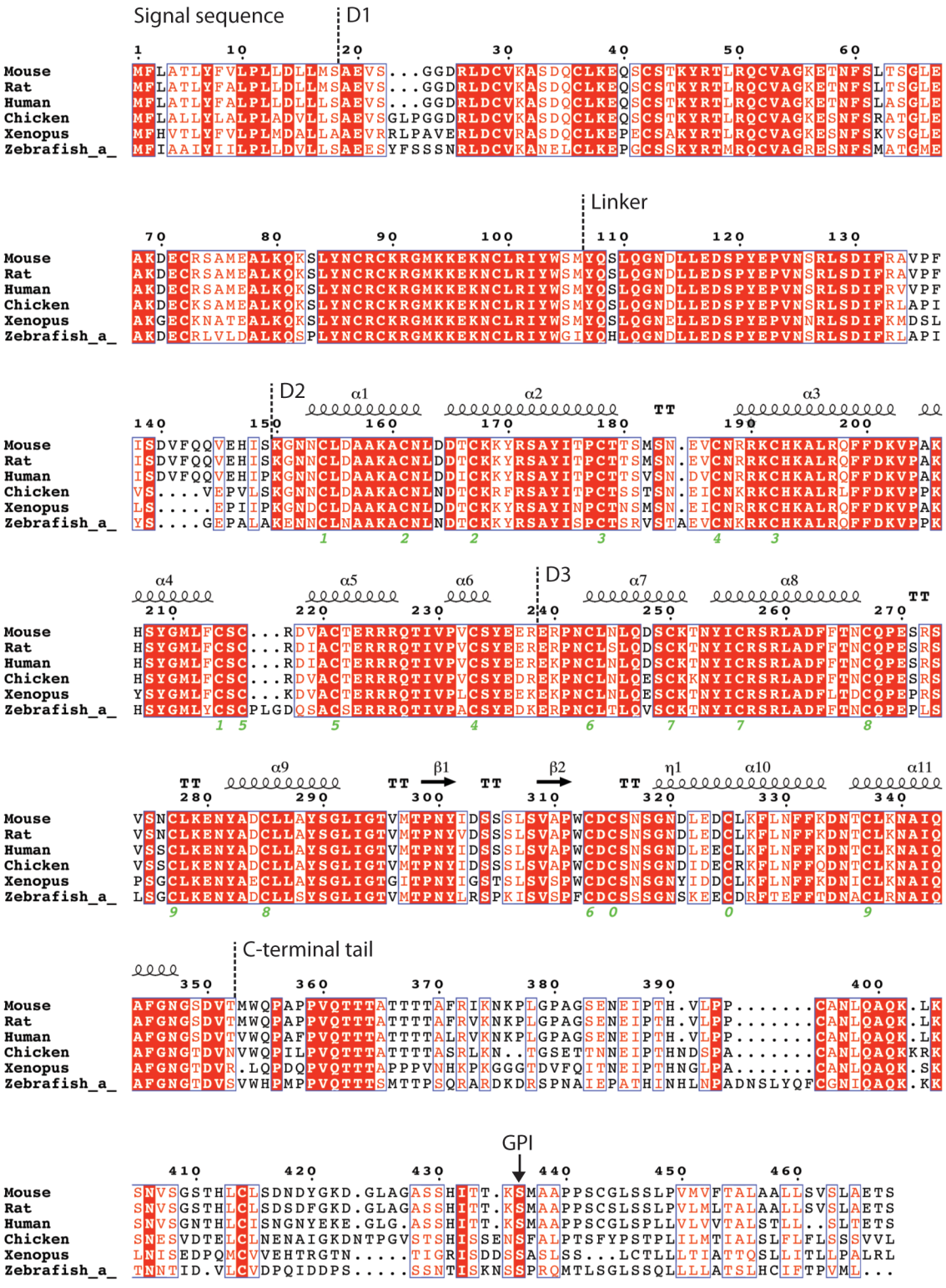



**Figure 4.1 Alignment of vertebrate GFRα1 amino acid sequences**

CLUSTALW2 alignment of GFRα1 amino acid sequences from three higher and three lower vertebrates: mouse (*Mus muscularis*), rat (*rattus norvegicus*), human (*homo sapien*), chicken (*Gallus gallus*), *Xenopus* (*Xenopus tropicalus*) and zebrafish (*Danio rerio*) isoform a. The secondary structure elements from the rat GFRα1(D2D3) crystal

structure (PDB 2V5E) are annotated above the alignment. Disulphide-linked pairs of cysteine residues present in the crystal structure are marked with green numbers beneath the alignment. Completely conserved residues are boxed and highlighted in red with white text, and similar residues boxed with red text. Dashed black lines indicate domain boundaries.

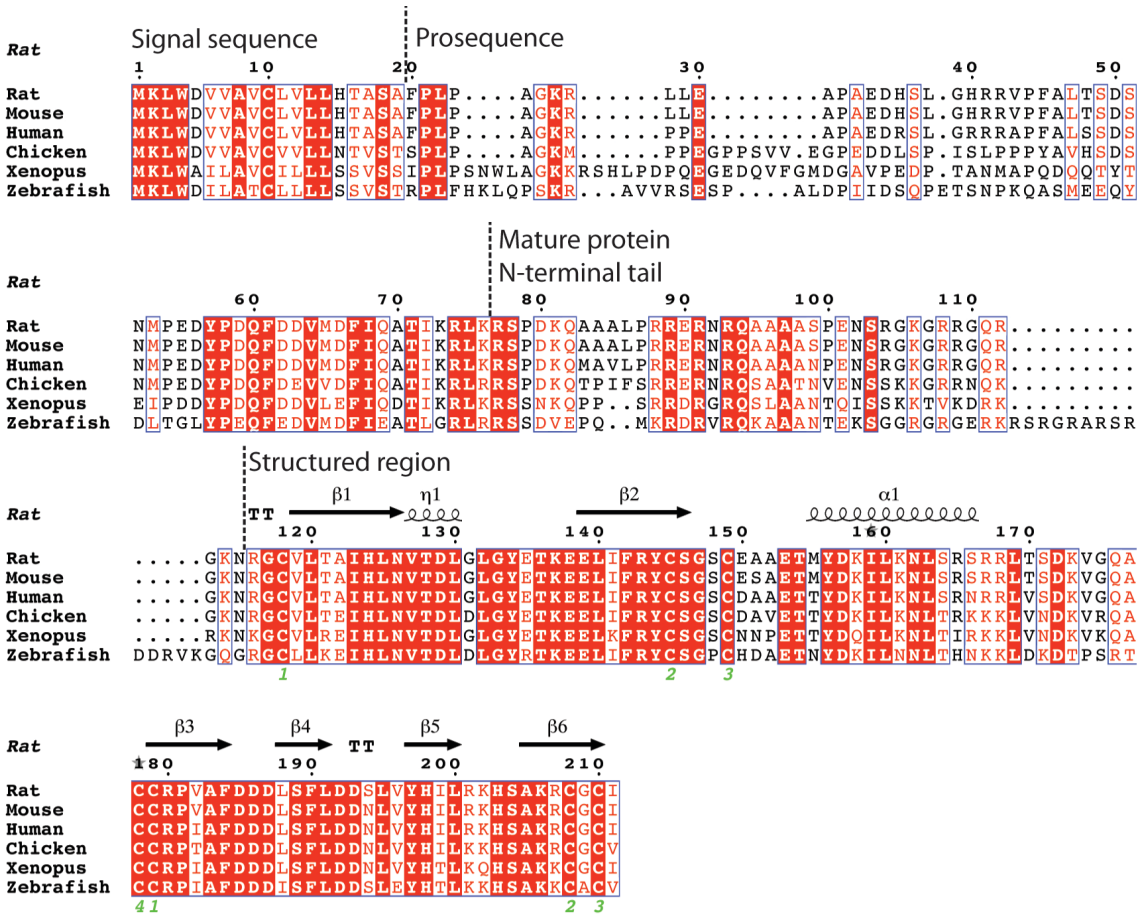



**Figure 4.2 Alignment of vertebrate GDNF amino acid sequences**

CLUSTALW2 alignment of GDNF amino acid sequences from three higher and three lower vertebrates. The secondary structure elements present in the human GDNF crystal structure (PDB 2V5E) are indicated above the alignment. The alignment is annotated as described for the GFRα1 alignment above (Figure 4.1). Dashed black lines indicate domain boundaries.

## 4.3 Reconstitution of the zebrafish ternary complex

### 4.3.1 Production of ligand/co-receptor

In order to generate a zebrafish RET ternary complex (zTC) for structural analysis it was necessary to develop a method of producing a RET ligand and corresponding co-

receptor. Given the availability of the clones and the direct comparison to the mTC I worked on producing zebrafish GDNF and GFRα1a (DNA courtesy of Iain Shepherd, Emory University). The mammalian GDNF/GFRα1 and Artemin/GFRα3 complexes have been produced using the baculovirus system, with sufficient yields for crystallisation screening (Parkash *et al.*, 2008; Wang *et al.*, 2006), and since I already had this expression system set up for the zRET-ECD production this was the clear starting point.

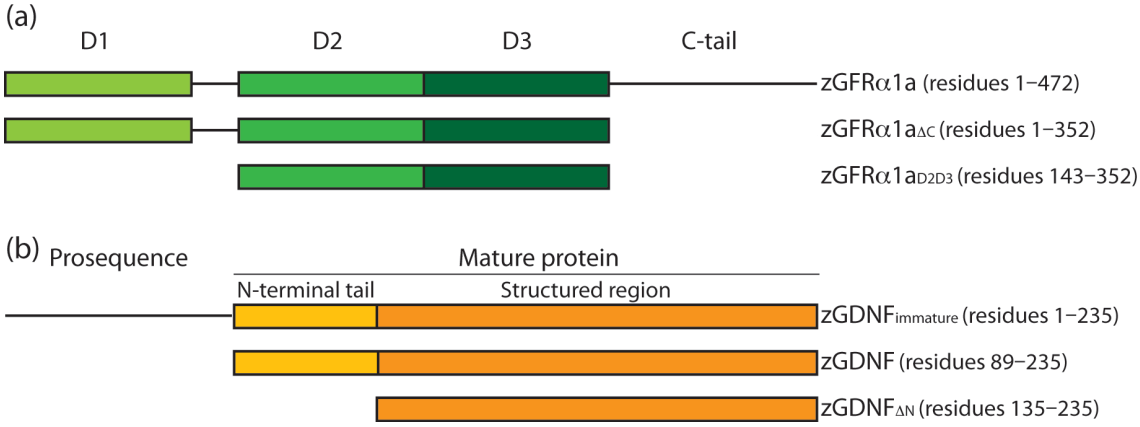



**Figure 4.3 Zebrafish ligand and co-receptor constructs examined in this thesis**
(a) zGFRα1a and (b) zGDNF constructs used in this thesis.

#### *4.3.1.1 Co-receptor construct design and protein expression*

GFRα1a consists of three discrete domains, designated D1 – D3, and an unstructured C-tail containing the GPI site. Using the crystal structure of mammalian GDNF/GFRα1 as a guide for the GFRα1 domain boundaries (Parkash *et al.*, 2008), I produced a zGFRα1a C-terminal truncated construct containing just the structured D1–D3 domains (zGFRα1a$_{\Delta C}$; residues 1–352; Figure 4.3a). I used the same expression cassette in the pBacPAK transfer vector as the zRET-ECD, with the native (zGFRα1a) signal sequence and cleavable C-terminal Protein A tag (section 2.3.2). Protein expression from the resulting baculovirus in Hi5 cells produced around 1 mg zGFRα1a$_{\Delta C}$ per litre cell culture. When expressed alone the protein tends to spontaneously proteolyse within 24 hours at 4 °C into two fragments: D1 alone and D2D3 (Figure 4.4b). This was not observed when co-expressed with GDNF suggesting the ligand protects the co-receptor from degradation to some extent.

#### *4.3.1.2 Ligand construct design and protein expression*

Since zGDNF is normally produced in the cell as a pro-protein, which is subsequently processed proteolytically, I designed a construct containing just the mature portion of protein (residues 89–235). This involved truncating the N-terminus of the gene, thus removing the native signal sequence. I therefore used a baculovirus transfer vector containing a honeybee melittin signal sequence (Tessier *et al.*, 1991). I left the protein untagged and co-expressed it with zGFRα1a$_{\Delta C}$. The yield of zGDNF however was very low, only detectable by western blotting following affinity purification of its binding partner zGFRα1a$_{\Delta C}$. This was the case even when infecting the Hi5 cells with the zGDNF baculovirus at an MOI of 5 and reducing the expression temperature to 18 °C.

In order to try to increase the protein expression I removed the unstructured N-terminus, which was disordered in the GDNF crystal structures (Eigenbrot and Gerber, 1997; Parkash *et al.*, 2008) and is not required for RET binding and activation (Baloh *et al.*, 2000). This construct defined as zGDNF$_{\Delta N}$, corresponding to residues 135–235, led to an improved yield of zGDNF$_{\Delta N}$ when co-expressed with zGFRα1a$_{\Delta C}$. Varying the ratio of the two viruses allowed similar levels of zGDNF$_{\Delta N}$ and zGFRα1a$_{\Delta C}$ to be obtained, with an overall protein yield of around 1 mg per litre (Figure 4.4a).

Non-reducing SDS-PAGE of the purified zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$ showed that the zGDNF$_{\Delta N}$ is mostly monomeric rather than a disulphide-linked dimer directly after affinity purification. During subsequent preparation of the ternary complex with zRET-ECD some of the zGDNF forms disulphide-linked dimers. The extent of dimerisation observed appears to be pH dependent, with zGDNF$_{\Delta N}$ remaining monomeric at pH 6.5, whereas at pH 7 some dimerisation occurs.

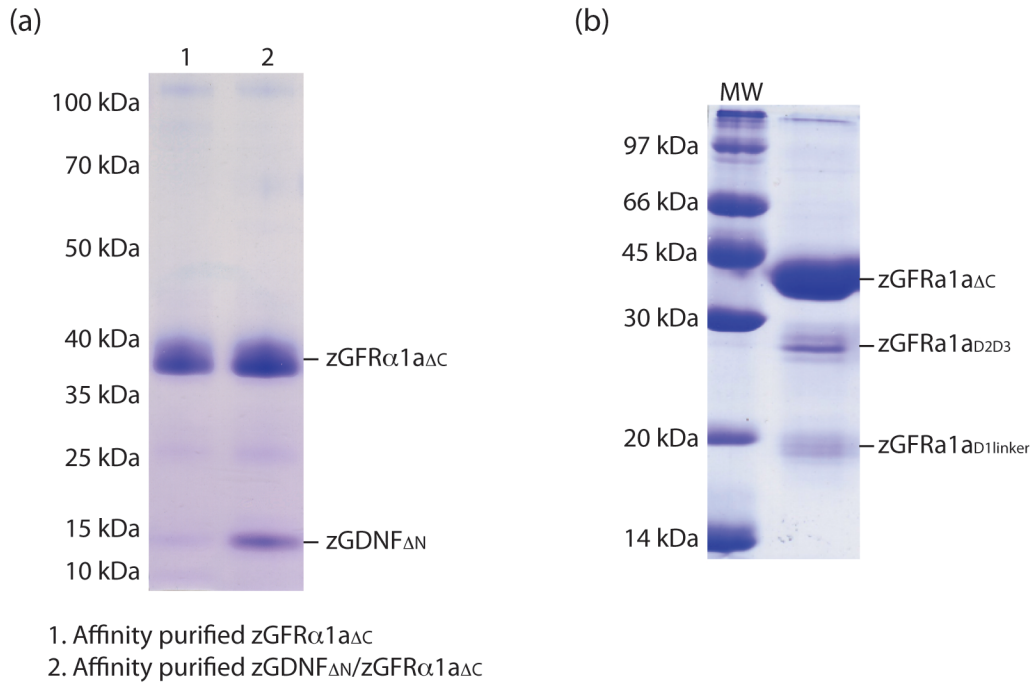



**Figure 4.4 Ligand and co-receptor production**

(a) Non-reducing SDS PAGE of affinity purified zGFRα1a$_{\Delta C}$ expressed alone (lane 1), and with zGDNF$_{\Delta N}$ (lane 2). (b) Non-reducing SDS PAGE of affinity purified zGFRα1a$_{\Delta C}$ alone after storage of the protein at 4 °C for 24 hours. The protein shows spontaneous proteolysis over time, with the identity of the fragments, as annotated, determined by mass spectrometry.

### 4.3.2 Assembly and calcium dependence of the zTC

To confirm the insect cell expressed proteins are functional and interact to form a RET/ligand/co-receptor ternary complex I performed a pull down assay. Briefly, the zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$/zRET-ECD ternary complex (zTC) was assembled on immobilised Protein A tagged zGFRα1a$_{\Delta C}$ as follows. Affinity purified zRET-ECD was added to zGFRα1a$_{\Delta C}$ attached to IgG-sepharose via its C-terminal Protein A tag in the presence or absence of both zGDNF$_{\Delta N}$ and calcium (2 mM $CaCl_2$). zRET-ECD binding to zGFRα1a$_{\Delta C}$ required the presence of both ligand and calcium (Figure 4.5). zRET-CLD(1–4) did not bind zGFRα1a$_{\Delta C}$ under these conditions, confirming the need for the CRD for ligand/co-receptor binding. It has not yet been tested whether the zRET-CRD can bind zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$ alone, although evidence from RET-ECD chimera studies suggests that the CRD is not sufficient for GDNF/GFRα binding (Kjaer and Ibanez, 2003a).

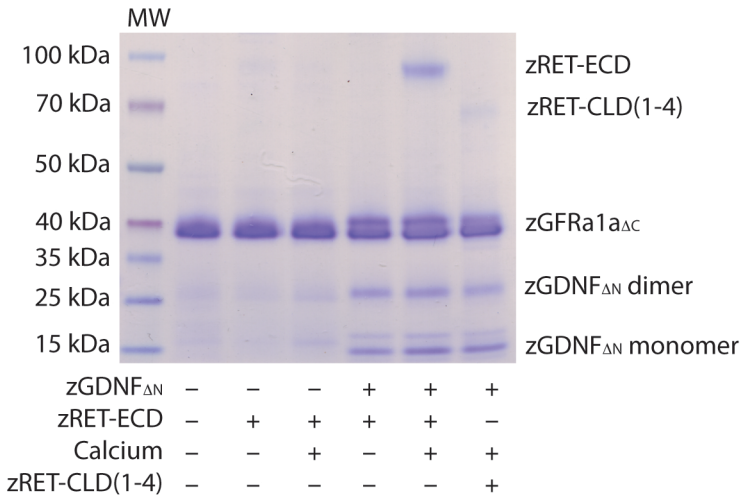



**Figure 4.5 Assembly of the zTC**

Non-reducing SDS PAGE of samples from zTC assembly on immobilised zGFRα1a$_{\Delta C}$. Ternary complex components and calcium (2 mM $CaCl_2$), as detailed below the gel, were added to the immobilised zGFRα1a$_{\Delta C}$ and allowed to bind for 1 hour on ice. Unbound protein was washed off, and the zGFRα1a$_{\Delta C}$ with bound proteins was eluted from the IgG-sepharose resin by 3C-protease. Eluted proteins were identified by SDS-PAGE, stained with Coomassie.

### 4.3.3 Purification of the zTC for structural analysis

The assembled zTC complex was eluted from the IgG-sepharose by 3C cleavage of the Protein A tag on zGFRα1a$_{\Delta C}$. The complex was then further purified by both size exclusion chromatography and anion exchange chromatography (Figure 4.6). The resulting complex appears stoichiometric by SDS-PAGE (bearing in mind the differing levels of Coomassie staining of the larger versus smaller proteins), is homogeneous and is stable in buffers containing up to at least 250 mM NaCl. The yield of zTC after size exclusion of complex assembled from a 2l culture of zRET-ECD and 5.5l culture of coexpressed zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$ is ~3 mg. Even better yields can be achieved by co-expressing all three proteins in Hi5 cells. However, since both the zRET-ECD and zGFRα1a$_{\Delta C}$ constructs carry the same Protein A tag, this method allows for less control over complex assembly/quality—although the complex can be separated from the constituent proteins by size exclusion chromatography. These yields make the zTC, unlike the mTC, a tractable target for crystallisation studies.

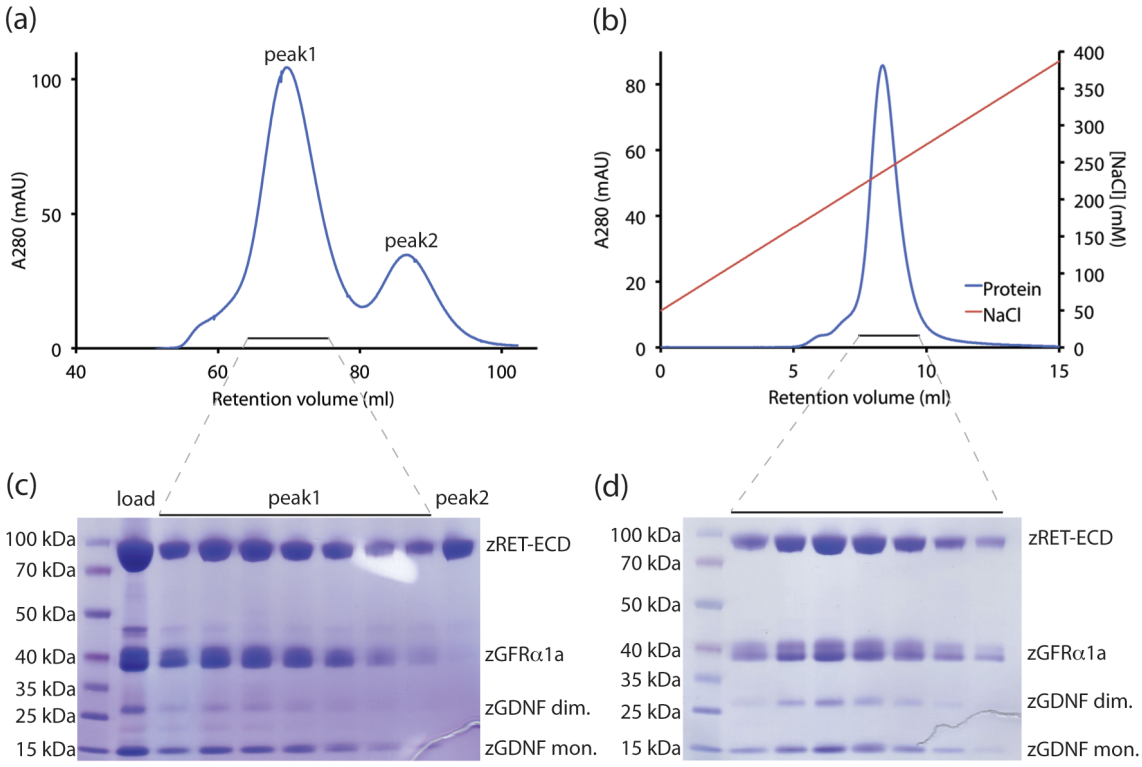



**Figure 4.6 zTC purification**

(a) Size exclusion chromatography of the zTC using a Superdex 200 column. (b) Anion exchange chromatography of the zTC, using a HiTrap Q column, following size exclusion. (c) Non-reducing SDS-PAGE of size exclusion and (d) anion exchange fractions.

### 4.3.4 Deglycosylation of the zTC

There are two predicted N-linked glycosylation sites in zGDNF, both of which reside within the structured region of the mature protein (Table 4.1). zGFRα1a has three predicted N-linked glycosylation sites, although all three are predicted with low confidence (Table 4.1). The N62 site is conserved and located in D1, N163 is not conserved and is in D2, and N350 is conserved and is at the very C-terminus of D3.

Deglycosylation of the zTC with Endo-F1 produced a smeared zRET-ECD band, suggesting that some of the N-linked glycans are protected by formation of the ternary complex (Figure 4.7). The upper band of zGFRα1a (see Figure 4.6d) is lost upon Endo-F1 treatment, suggesting it corresponds to a partially occupied glycosylation site.

| Position | Sequence | Potential | Jury agreement | N-Glyc result |
|---|---|---|---|---|
| ***zGDNF*** | | | | |
| 150 | NVTD | 0.7628 | (9/9) | +++ |
| 186 | NLTH | 0.6570 | (9/9) | ++ |
| ***zGFRα1a*** | | | | |
| 62 | NFSM | 0.5458 | (7/9) | + |
| 163 | NDTC | 0.5309 | (5/9) | + |
| 350 | NGTD | 0.6290 | (7/9) | + |
| 417 | NNTI | 0.4877 | (4/9) | - |
| 441 | NSSP | 0.1052 | (9/9) | --- |

**Table 4.1 Predicted N-linked glycosylation sites in zGDNF and zGFRα1a**

Analysis of all the NXS/T N-linked glycosylation consensus sites in zGDNF and zGFRα1a using the NetNGlyc server in ExPASY, which assesses whether the site is likely to be glycosylated on the basis of the primary amino acid sequence. Position = sequence position of the asparagine under consideration. Potential = the likelihood of glycosylation based on the averaged probability results of nine neural networks. A predicted site has a potential of > 0.5. Jury agreement = the number of the neural networks that support the prediction. N-Glyc result = indicates the prediction and the confidence level of the prediction: + = predicted glycosylated; - = predicted not glycosylated. The number of +/- indicates relative confidence in the prediction.

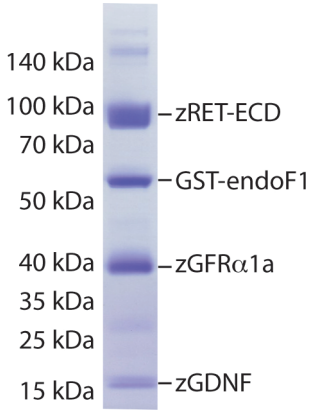



**Figure 4.7 Deglycosylation of assembled zTC with Endo-F1**

Non-reducing SDS PAGE of purified zTC after incubation with Endo-F1 for 20 hours at 4 °C.

### 4.3.5 zTC instability

The purified zTC was initially thought to be reasonably stable since it can be separated from the excess constituent proteins by size exclusion, is stable in anion exchange chromatography, and if a second size exclusion is carried out the complex elutes as a single peak. The size exclusion purified complex also gave a single peak in DLS, with

an approximate MW of 480–500 kDa. However, when the size exclusion and anion exchange chromatography purified zTC complex was visualised by negative stain EM, intact complexes of a similar size to those seen in the mTC were few: The grids were dominated by much smaller particles, consistent with individual components of the zTC and indicating that the complex is not stable under negative stain EM conditions.

Subsequent experiments to look into this apparent instability revealed that the purified complex consisted of a mixture of different sub-complexes. Native PAGE showed four bands (Figure 4.8b). The smallest, running at 150 kDa, corresponds to where zRET-ECD alone runs, although peptides from all three proteins could be detected by mass spectrometry of this and all the other bands. The other three bands correspond to MWs of approximately 500 kDa, 360 kDa and 250 kDa. Given that zRET-ECD runs much larger than its actual MW by native PAGE (even considering the glycosylation) it is tempting to speculate that the top band corresponds to the 2RET:2GFRα1:2GDNF complex, with the second band corresponding to the loss of a RET, and the third either both RETs or, more consistently with the relative MWs, a RET and a GFRα1. This hypothesis would fit with the proposal of (Schlee *et al.*, 2006); that the ternary complex assembles *in vivo* via 1RET:1GFRα1:2GDNF and subsequent 1RET:2GFRα1:2GDNF intermediates. A SEC-MALS experiment with the purified zTC complex also showed that the complex sample is not homogeneous, with the calculated MW of the eluted material decreasing from 420 kDa to 200 kDa (Figure 4.8).

This instability of the recombinant zTC is not seen for the mTC. This could be due to the zTC simply being less stable than the mTC, or a sub-optimal buffer composition for the zTC. It could also be due to the recombinant protein production methods: The zTC components were produced in insect cells, which means the glycans added to the proteins will not be precisely the same as those added natively in the zebrafish. This may have an impact on the complex if any of the glycans play a role in complex assembly/stability. In the mTC, the hRET-ECD and rGFRα1 were both produced in mammalian cells, and therefore their glycosylation is likely to be closer to native. Secondly the zTC $zGDNF_{\Delta N}$ and $zGFR\alpha1a_{\Delta C}$ are not the full-length proteins but had predicted unstructured elements removed during cloning, whereas the components of the mTC are all full length, aside from the GPI-link site of rGFRα1. It may be that the

N-terminal tail of zGDNF, which has been shown by crosslinking to interact with RET (Amoresano *et al.*, 2005), or the C-tail of zGFRα1 are important for complex stability.

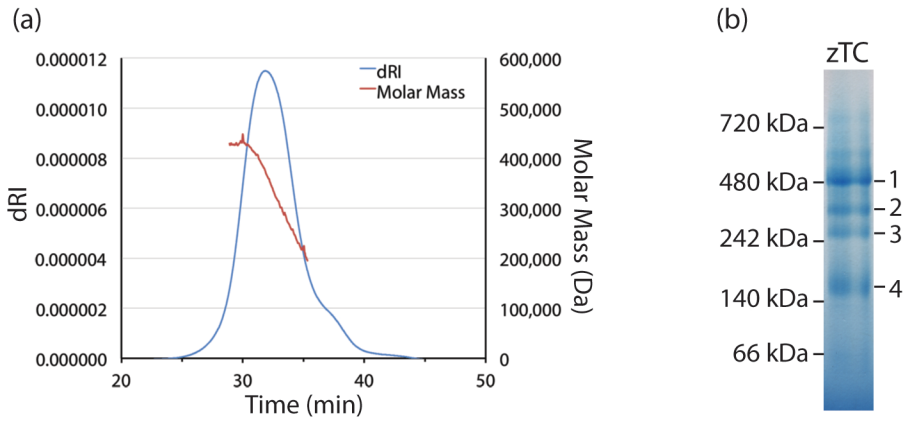



**Figure 4.8 zTC instability**

(a) SEC-MALS results, using a Superose 6 column, of size exclusion purified zTC. dRI = differential refractive index. (b) Coomassie stained native PAGE of size exclusion purified zTC.

## 4.4 SAXS analysis of the zebrafish ternary complex

The heterogeneity of the purified zTC means that it is not a good sample for SAXS analysis, since SAXS requires a homogeneous sample for molecular shape analysis of samples of unknown structure. The decreasing complex size across the size exclusion peak as seen in SEC-MALS (Figure 4.8) is reflected in the decreasing radius of gyration (Rg) observed across the HPLC peak in an online-HPLC SAXS experiment (Figure 4.9). The Rg decreases from 61.7–59 Å across the HPLC peak.

Analysis of averaged SAXS curves from small sections across the HPLC peak, corresponding to similar Rg's, does result in apparently reasonable pair distance distribution function, based on analysis of the data with GNOM. However, as expected given the known heterogeneity, no meaningful results can be produced with *ab initio* model generation from these data.

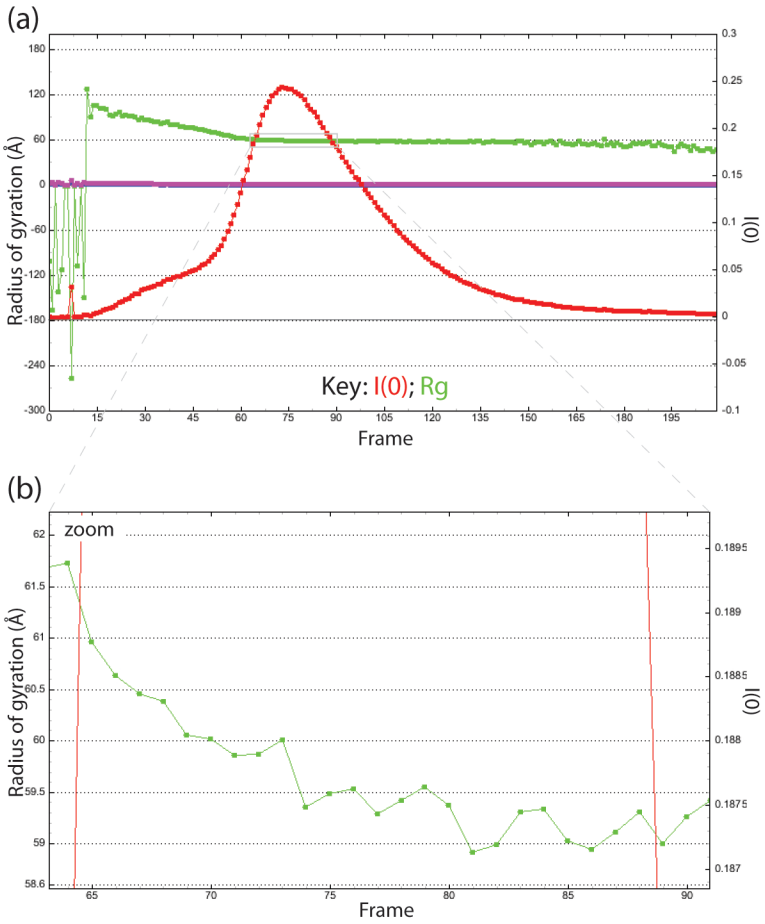



**Figure 4.9 Analysis of SAXS data with online HPLC of zTC**

(a) Graph showing the radius of gyration (Rg) and the extrapolated intensity at s = 0 (I(0)) calculated, using the Guiner approximation, for each data collection. The I(0) indicates the protein concentration in the beam, and therefore represents the HPLC peak. (b) Close up of the variation in Rg across the HPLC elution peak profile.

## 4.5 Negative stain EM of the zebrafish ternary complex

### 4.5.1 Statement of contributions

The work described in this section (4.5) was a collaborative effort primarily between Dr. Fabienne Beuron (Imperial Cancer Research, London) and myself. I produced the zTC sample, as described above. Dr. Fabienne Beuron prepared the EM grids of the sample and collected the images. I picked particles, generated classes, and performed projection matching to generate a preliminary 3D reconstruction.

### 4.5.2 Preparation of cross-linked zTC for negative stain EM

As previously stated (section 4.3.4), negatively stained EM grids of the size exclusion and anion exchange purified zTC were dominated by smaller particles than seen for the mTC. This observation was supported by the native PAGE and SEC-MALS data, showing the zTC sample to consist of a mixture of different sub-complexes. This untreated sample was therefore not useful for EM analysis of the complete complex. However, crosslinking of EM samples has been shown to stabilise labile biological samples on occasion (Kastner *et al.*, 2008). Indeed we found that addition of glutaraldehyde to a final concentration of 0.05% for 3 minutes resulted in a homogeneous sample with particles of similar size to the mTC (Figure 4.10). The EM grids were negatively stained with 2% (w/v) uranyl acetate and images were collected on a Tecnai T20 electron microscope, equipped with a 4000×4000 pixel CCD detector, at 80000× magnification as for the mTC (see section 3.2.4).

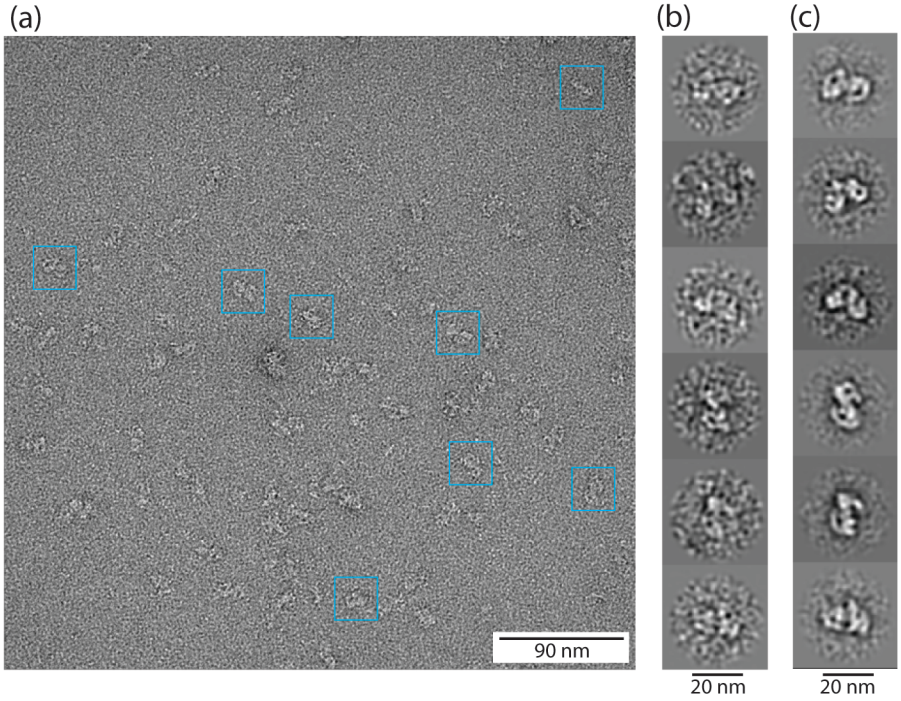



**Figure 4.10 Negative stain EM of zTC**

(a) Example image of the cross-linked zTC EM grid, negatively stained with 2% (w/v) uranyl acetate. (b) Examples of picked particle images after filtering and normalisation. (c) Corresponding class averages from MSA-based classification following particle alignment to the mTC 3D reconstruction.

### 4.5.3 Particle picking and generation of class averages

From the collected images I picked 7500 particle images, which were filtered and normalised as for the mTC (see section 3.2.5). The particle image stack was dominated by the 'figure of eight' view (first image in Figure 4.10c), which was one of the characteristic views of the mTC. The other defining view of the mTC was the 'leg' view, which occurs with far less frequency in the zTC images, and appears altered. In particular 'leg' views with clear base group regions and triangularly splayed 'legs' do not appear to be present in the current zTC stack, although it may simply be that it is underrepresented in the zTC sample on the EM grid. The views most similar to the mTC 'leg' view show more of a squared top, at the point where the legs come together (second and third image in Figure 4.10c).

The dominance of the 'figure of eight' view in the zTC particle image stack meant class assignment based on reference-free alignment using EMAN (Ludtke *et al.*, 1999) resulted in very few non-'figure of eight' class averages (Figure 4.11). Aligning the stack to forward projections of the mTC 3D reconstruction in SPIDER (Frank *et al.*, 1996) followed by MSA-based class assignment in IMAGIC (van Heel *et al.*, 1996), resulted in a much better classification. The class averages corresponding to the 'figure-of-eight' view are still the most abundant, but other views, such as the 'leg'-type view described above are also represented (Figure 4.10b&c).

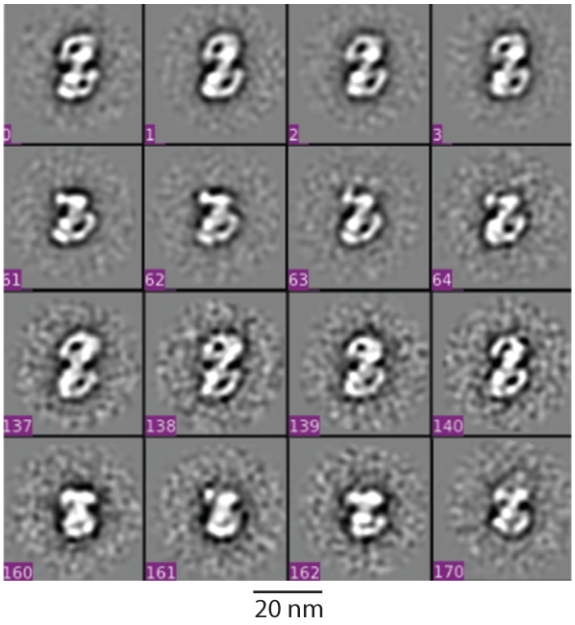



**Figure 4.11 zTC representative class averages from reference-free alignment and classification**

A selection of zTC class averages from the 300 generated by reference-free alignment and classification using EMAN from the 7500 particle image stack.

### 4.5.4 Projection matching to the mammalian ternary complex reconstruction

Given the similarity between the classes from the mTC and the zTC, I tried to see if I could generate a reasonable 3D reconstruction of the zTC by starting with projection matching to the mTC using IMAGIC. The results of an initial round of projection matching to the mTC are shown in Figure 4.12. The zTC classes used in this projection matching procedure were generated following alignment of the zTC particle image stack to mTC re-projections. Comparison between the zTC classes and the matched mTC re-projections showed a good correspondence for around half of the input zTC classes. The map showing the Euler angles assigned to these classes (Figure 4.12c) further highlights the dominance of the 'figure-of-eight' view, which corresponds to the cluster around (0,0) and (180,0). Expansion of the particle image stack and further improvements in classification will be needed to improve the coverage of the Euler sphere and to lead to a final 3D reconstruction of the zTC.

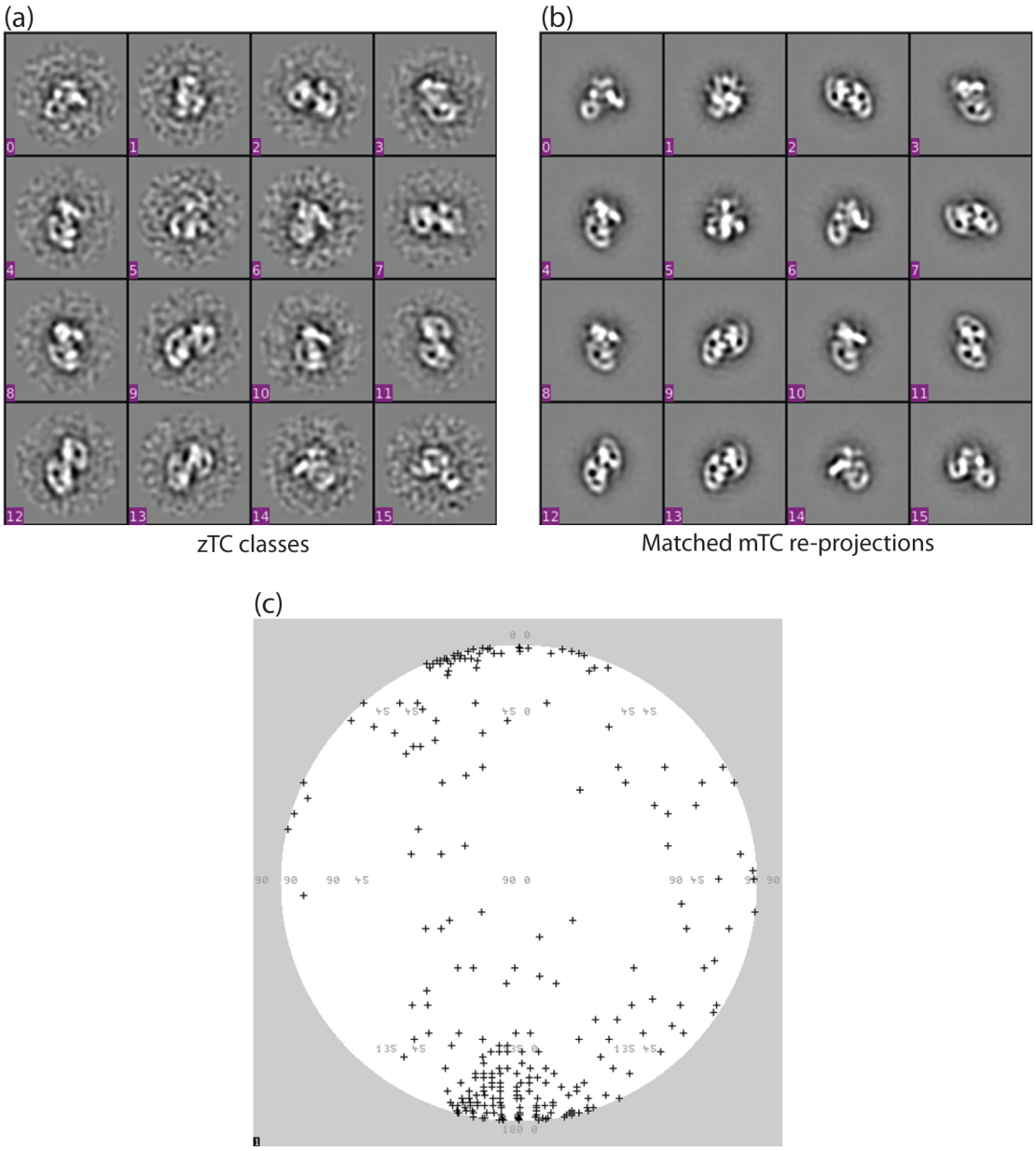



**Figure 4.12 Towards a zTC 3D reconstruction**

Projection matching of the zTC classes to re-projections of the mTC 3D reconstruction. (a) A selection of zTC class averages. (b) Corresponding re-projections from the mTC 3D reconstruction. (c) Map of Euler angles assigned to the zTC classes showing a good match to an mTC re-projection.

## 4.6 Conclusions

In this chapter I have presented a method for the production of the zebrafish RET/GDNF/GFRα1a ternary complex using the recombinant baculoviruses to infect an insect cells host for protein production. The protein yields and purity are sufficiently

high to consider crystallisation screening. However, the recombinant zTC is unstable, with the purified sample made up of a mixture of the complete 2:2:2 complex and different sub-complexes when subjected to native PAGE or observed by EM following staining with 2% (w/v) uranyl acetate. Low level cross-linking with glutaldehyde was sufficient to produce a visually homogeneous sample in negative stain EM. Further buffer screening may lead to increased complex stability allowing the zTC to then be used confidently in crystallisation screening.

Analysis of the cross-linked zTC negative stain EM images revealed the zTC looks very similar to the mTC, suggesting the overall arrangement of the complex components is the same. In particular, the unchanged nature of the 'figure-of-eight' view would indicate that the arrangement of RET is unchanged, assuming the interpretation of the mTC EM 3D reconstruction presented in the previous chapter is correct, and RET is indeed located in the wings of the 'figure-of-eight'. There are some hints of differences between the complexes from the class averages, particularly in the 'base' region of the 'leg' view, although these could simply be a reflection of different preferred orientations of the complexes on the EM grid. Increasing the size of the particle image stack to allow production of a 3D reconstruction of the zTC will be required to assess the impact of these apparent differences fully. It should be noted that the zGFRα1a$_{\Delta C}$ construct lacks the C-tail, which is present in the rGFRα1 within the mTC. Since this rGFRα1 C-tail may be contributing to the observed density in the 'base' region of the mTC 3D reconstruction, one might expect a reduction in the size of the 'base' region in the zTC.

Finally pull-down assays confirmed that zRET-ECD, like hRET-ECD (Anders *et al.*, 2001), requires calcium for binding the ligand/co-receptor, consistent with the conservation of the calcium-binding residues at the CLD2–CLD3 junction. Furthermore, I have also shown that the CRD is required for zRET binding to ligand/co-receptor supporting a key role for this domain in ligand/co-receptor engagement.

# Chapter 5. RET tyrosine kinase domain regulation and chemical inhibition

## 5.1 Chapter aims

This chapter describes efforts to crystallise the RET tyrosine kinase domain (RET-KD) with different ligands and inhibitors bound within the RET nucleotide-binding pocket to characterise the interaction of these small molecules with the RET-KD. I present the crystal structures of the RET-KD bound to the clinically relevant RET inhibitor sunitinib, whose mode of binding to RET was unknown; and a high-resolution structure of the RET-KD bound to adenosine, which revealed some new insights into the RET-KD structure and flexibility.

In addition, efforts were made to obtain a different crystal form of the RET-KD without the dominant head-to-tail dimer observed in all other RET-KD crystal forms and to attempt to capture novel RET-KD conformations that may give clues to *cis*- or *trans*-inhibitory mechanisms operating within RET. Previous RET-KD crystal structures identified a *trans*-inhibited dimer of the RET-KD, whose physiological relevance remained to be proven. The significant size of the protein interaction interface within the *trans*-inhibited dimer and the observation that several interface residues were MEN2 oncogenic mutation sites, including the aggressive M918T mutation, suggested a closer look may be important (Knowles *et al.*, 2006). The longer-term goal in understanding RET regulation was to be able to interpret how oncogenic RET-KD mutations subvert RET regulation leading to tumourigenesis.

At the start of this thesis, no structure was yet available for the RET-KD bound to either ATP (or an analogue) or a peptide substrate. Therefore a further aim of this chapter was to determine a RET-KD structure bound to an ATP-analogue and a tyrosine-containing peptide substrate, since it was hoped this would allow us characterise the RET-KD fully poised for catalysis as well as the interactions with ATP and substrate peptide. I present successful crystallisation of such a complex, although further optimisation of these crystals is required to improve the resolution of X-ray diffraction data and allow the structure to be solved by molecular replacement.

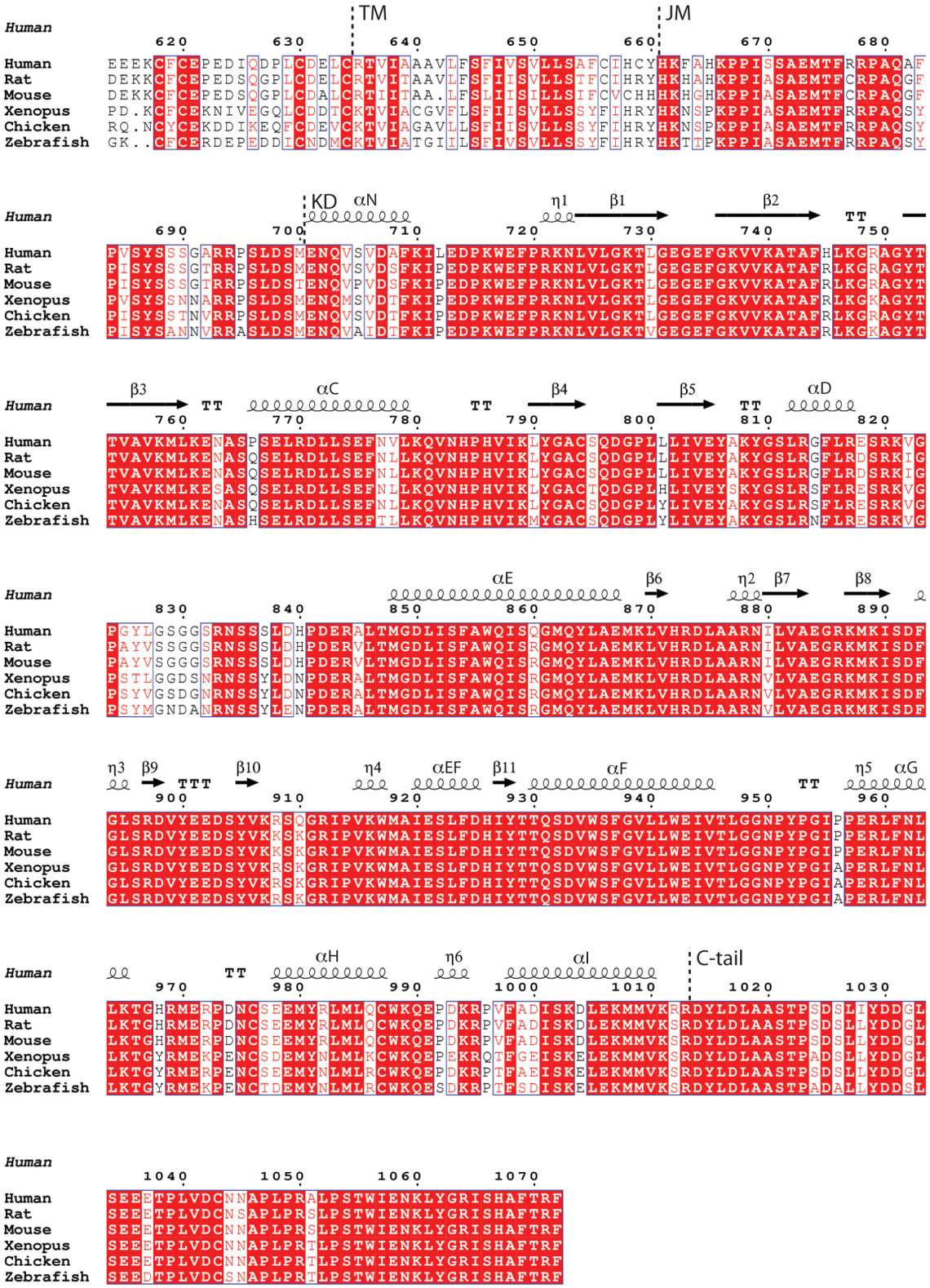



**Figure 5.1 Alignment of vertebrate RET-ICD amino acid sequences**

CLUSTALW2 alignment of 3 higher and 3 lower vertebrate RET-ICDs: human (*homo sapien*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), *Xenopus* (*Xenopus laevis*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*). Completely conserved residues are

boxed and highlighted in red with white text, and similar residues boxed with red text. Secondary structure elements observed in the human RET-KD crystal structure are annotated above the alignment and domain boundaries are indicated by dashed black lines.

## 5.2 Protein production

### 5.2.1 Sequence alignment of RET intracellular domains

Alignment of RET-ICD amino acid sequences from three mammalian and three lower vertebrate RETs shows very high sequence conservation throughout the ICD, with 85% sequence identity between human and zebrafish (Figure 5.1). Within the KD the most amino acid sequence divergence occurs in the glycine and serine rich kinase insert region (residues 820–845)

### 5.2.2 Protein expression and purification

Tyrosine kinases have been difficult to produce in *E. coli*, although newer approaches of co-expressing tyrosine phosphatases alleviate some of the toxicity of overexpressing tyrosine kinases in bacteria (Seeliger *et al.*, 2005). Nevertheless, a detailed protocol was available in the laboratory to produce recombinant human RET-KD (hRET-KD; residues 705–1013) using an insect cell expression host and a recombinant baculovirus (Knowles *et al.*, 2006). This protocol was robust and gave reasonable yields (~1 mg/l cell culture) of monodisperse protein. In order to improve yields further we purchased a hRET-KD gene codon-optimised for insect cell expression from GeneART, which I subcloned into the same baculovirus transfer vector used previously for hRET-KD baculovirus production. Use of the codon-optimised construct improved yields 4× with the same protein production protocol.

Briefly hRET-KD, excluding the highly flexible kinase insert (residues 827–840), was expressed with a 3C-protease cleavable N-terminal glutathione S-transferase (GST) affinity tag, in both Hi5 and Sf9 cells. The single step affinity-purified protein was judged to be pure by SDS-PAGE (Figure 5.2a), and was known to crystallise without further purification. The RET-KD is non-phosphorylated as purified out of insect cells. To produce a single fully phosphorylated species, hRET-KD was phosphorylated by

addition of 5 mM ATP and 10 mM $MgCl_2$ for 6 hours on ice. The ATP and $MgCl_2$ were then removed by dialysis.

To characterise the phosphorylated hRET-KD protein (pRET-KD), size exclusion chromatography was carried out. Using a standard 100 mM NaCl concentration in the size exclusion buffer the majority of the pRET-KD is associated as an aggregate (Figure 5.2b). However at higher salt (300 mM NaCl), the proportion of monomer is greater (Figure 5.2c). This result is reflected in the long concentration times required to obtain pRET-KD concentrations of > 3 mg/ml, particularly in the absence of nucleotide/inhibitor binding in the active site.

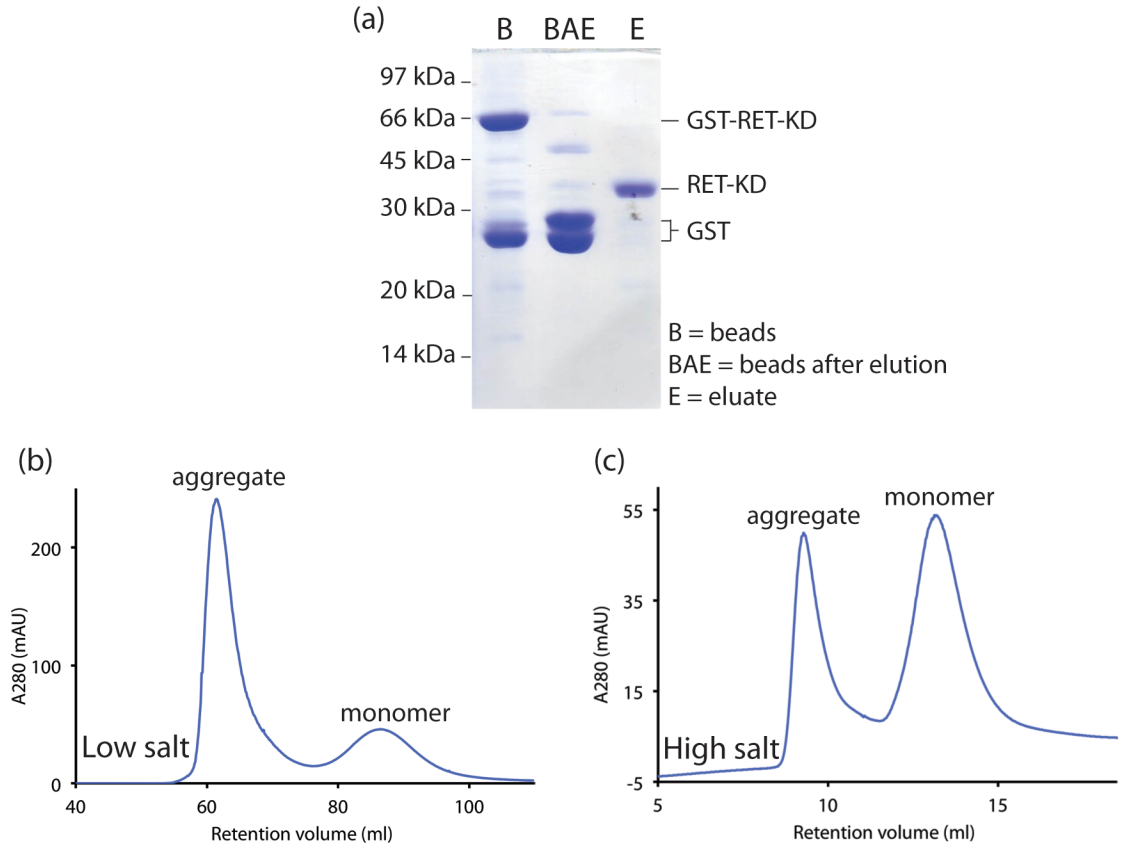



**Figure 5.2 pRET-KD purification**

(a) Reducing SDS-PAGE of RET-KD affinity purification samples. GST-RET-KD was affinity purified from cell lysates using glutathione-sepharose resin, followed by elution of RET-KD using 3C-protease. (b) & (c) Size exclusion chromatography of pRET-KD, using Superdex 75 columns, with (b) 100mM NaCl, or (c) 300 mM NaCl in the running buffer.

## 5.3 Structure of sunitinib bound to the RET-KD

All previous RET-KD structures bound to chemical inhibitors solved in the Structural Biology Laboratory show the compounds binding to an 'active' *DFG-in* conformation (Knowles *et al.*, 2006). One example is the RET-KD structure bound to vandetanib, a clinically relevant inhibitor that became the first drug approved by the FDA for use in MTC (Thornton *et al.*, 2012). However issues with the efficacy of vandetanib are known and discussed in Chapter 1 (section 1.8.2). Furthermore vandetanib exhibits no activity against RET gatekeeper mutations such as V804M, therefore resistance mutations to vandetanib would be expected at the gatekeeper. This would suggest a need for other potent clinically relevant RET inhibitors such as sorafenib and sunitinib (Sutent®) to treat resistance or as a combination therapy. These compounds have been suggested to bind an inactive *DFG-out* conformation (Chow and Eckhardt, 2007; Gajiwala *et al.*, 2009; Plaza-Menacho *et al.*, 2007). I therefore sought to crystallise the RET-KD with sunitinib bound to structurally characterise the interaction and determine whether sunitinib binding in the RET-KD ATP-cleft could trap a *DFG-out* conformation.

### 5.3.1 Crystallisation and crystal optimisation

To characterise structurally the binding of sunitinib to the RET-KD co-crystallisation experiments were carried out. To do this sunitinib (Toronto Research Chemicals) was dissolved in DMSO and added to pRET-KD at a 3:1 inhibitor:protein molar ratio. Following incubation for 1 hour at 4 °C to allow binding, the protein:inhibitor complex was concentrated to 3–5 mg/ml for crystal trials. Crystals were obtained in sodium formate and ammonium sulphate-based conditions previously identified for other RET-KD:inhibitor complexes (Knowles *et al.*, 2006). The ammonium sulphate crystals were hexagonal, whereas the sodium formate crystals were predominantly stacked square plates (Figure 5.3).

Crystal X-ray diffraction screening of these two crystal forms showed that the plate morphology crystals diffracted to a much higher resolution (up to 2.6 Å) than the hexagonal crystal form (up to 10 Å). Therefore I proceeded to further optimise the plate

crystals in order to produce single crystals and eliminate multiple lattices evident from the diffraction data.

Optimisation experiments involved varying the protein concentration, the protein:precipitant ratio in the crystallisation drop and performing an additive screen. Increasing the protein concentration from 3 to 4 mg/ml, accompanied by a slight reduction in the sodium formate concentration to 1.8–2 M resulted in some single crystals, as well as stacked plates (Figure 5.4a).

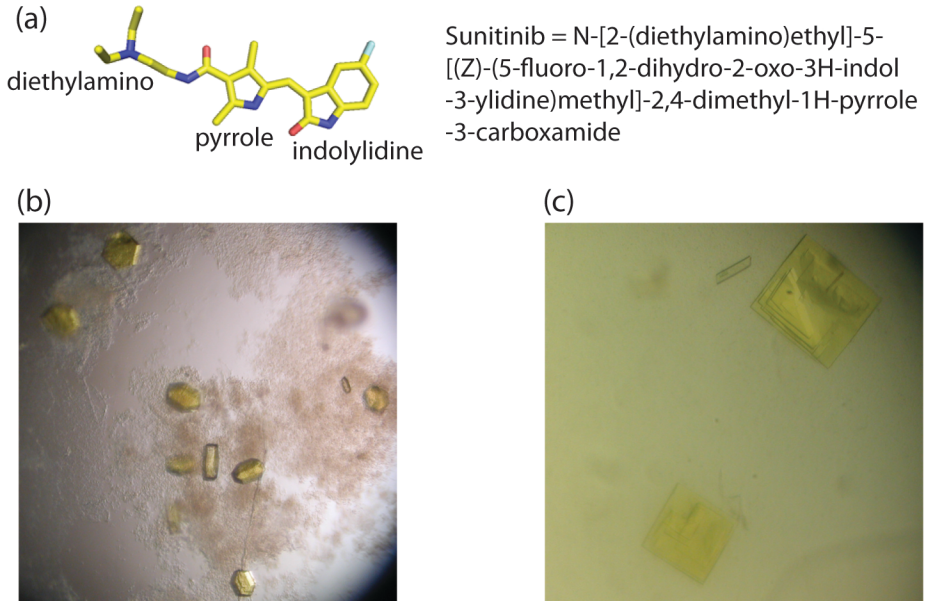



**Figure 5.3 Initial pRET-KD:sunitinib crystals**

(a) Chemical structure of sunitinib. (b) Hexagonal pRET-KD:sunitinib crystals grown in 0.8–1 M ammonium sulphate and 0.1 M sodium citrate (pH 5) with 5 mg/ml protein using 1:1 protein:precipitant drops. (c) Stacks of square plate crystals grown in 2.1 M sodium formate, 0.1 M sodium citrate (pH 5.5), with 3 mg/ml RET-KD using 3:2 protein:precipitant drops.

### 5.3.2 pRET-KD:sunitinib structure determination

Around 30 of the apparently single plate crystals, each around 80 µm long, were cryoprotected with n-paratone and flash-frozen in liquid nitrogen. These crystals were screened for diffraction at Diamond Light Source, and datasets were collected from those exhibiting better than 3 Å resolution. The best dataset obtained showed diffraction to Bragg spacings of 2 Å (Figure 5.4b&c). However, as indicated by the high $R_{merge}$ in

the outer shell (60.2%), the data quality was poor beyond 2.2 Å, and was therefore reprocessed to 2.2 Å (Table 5.1).

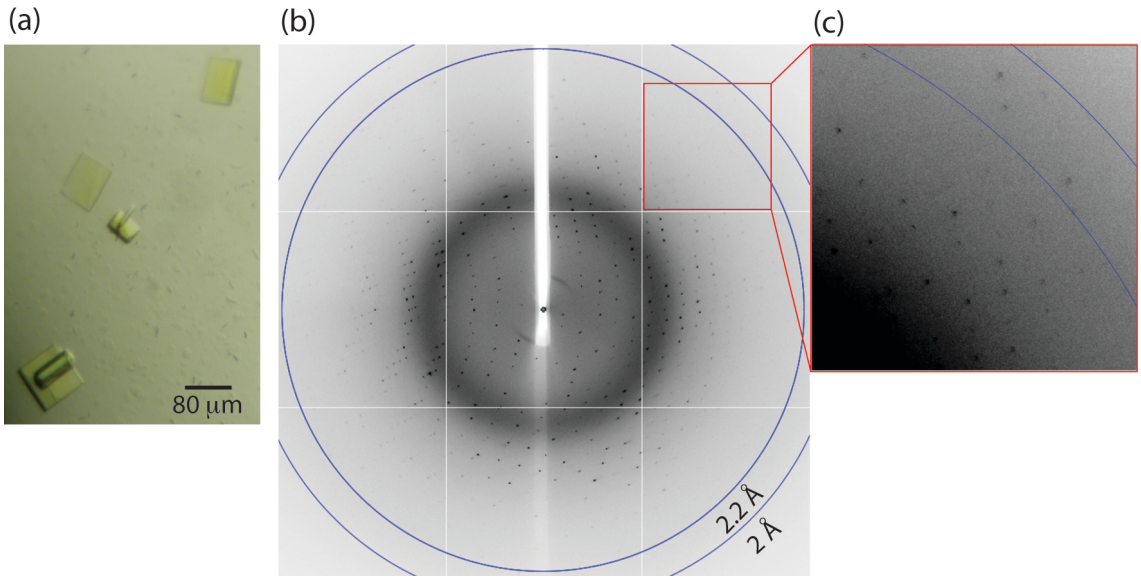



**Figure 5.4 pRET-KD:sunitinib optimised crystals and diffraction**

(a) Optimised pRET-KD:sunitinib crystals. (b) & (c) Diffraction image from the 2.2 Å dataset used in structure determination.

| Data collection | |
|---|---|
| **X-ray source** | Diamond Light Source I-03 |
| **Wavelength / Å** | 0.9800 |
| **Space group** | C2 |
| **Unit cell dimensions** | |
| **a, b, c / Å** | 73.32, 70.01, 78.97 |
| **α, β, γ / °** | 90.00, 101.54, 90.00 |
| **Resolution / Å** | 45.04–2.20 (2.32–2.20) |
| **Total reflections** | 53468 (7939) |
| **Unique reflections** | 19457 (2866) |
| **Multiplicity** | 2.7 (2.8) |
| **Completeness / %** | 97.7 (99.4) |
| **I/σ** | 10.1 (3.6) |
| **$R_{merge}$ / %** | 7.5 (32.4) |
| **$R_{p.i.m.}$ / %** | 5.3 (22.8) |
| **Wilson B-factor / Å²** | 28.7 |

**Table 5.1 Data collection statistics for pRET-KD:sunitinib**

Values for the outermost shell are shown in parenthesises.

#### 5.3.2.1 Structure determination by molecular replacement

The dataset was processed using MOSFLM (Powell, 1999) and scaled and merged in the CCP4 suite (Collaborative Computational Project, 1994). POINTLESS (Evans, 2006) confirmed the space group was C2, and cell content analysis indicated that 1 molecule was present in the asymmetric unit. The structure was solved by molecular replacement in PHASER (McCoy *et al.*, 2007), using the pRET-KD:PP1 structure (2IVV) as a model. PHASER identified a single solution with rotation function Z-score of 41.5 and translation function Z-score of 28.5. The final solution following automatic refinement in PHASER had a log-likelihood gain of 2611.5 and an $R_{factor}$ of 35.2%.

#### 5.3.2.2 Structure refinement with REFMAC

The pRET-KD:sunitinib structure was refined at 2.2 Å using REFMAC (Murshudov *et al.*, 1997) and model building was carried out in COOT (Emsley and Cowtan, 2004). Each round of refinement in REFMAC was carried out using ten rounds of TLS (translation, libration, screw rotation) refinement followed by ten rounds of restrained refinement. TLS groups used corresponded to the N and C-lobes (residues 716–808 and 812–1011), and this refinement procedure allowed separate rigid body refinement of these two relatively mobile subdomains.

The final refined model contains five N-terminal vector derived residues and residues 705–1011 from the RET-KD, excepting the 711–715 loop and 823–843 from the kinase insert (residues 827–840 were absent from the construct). It also contains one sunitinib molecule and two formate molecules per KD. The $R_{work}$ and $R_{free}$ are 19.8% and 26.0% respectively, and the structure has good geometry (see Table 5.2).

| Refinement | |
|---|---|
| **Resolution / Å** | 45.04–2.20 (2.26–2.20) |
| **Completeness / %** | 97.4 (99.3) |
| **No. of reflections** | 18467 (1366) |
| **$R_{work}$ / %** | 19.8 (20.4) |
| **$R_{free}$ / %** | 26.0 (27.0) |
| **Correlation coefficient (Fo-Fc) / %** | 93.9 |
| **Protein residues** | 279 |
| **Ligand molecules** | 3 |
| **Water molecules** | 52 |
| **Total non-hydrogen atoms** | 2246 |
| **Mean B-factor / Å** | 26.9 |
| **Protein** | 28.7 |
| **Sunitinib** | 28.1 |
| **Water** | 29.6 |
| **Ramachandran Plot / %** | |
| **Favoured** | 94.98 |
| **Allowed** | 3.58 |
| **Outliers** | 1.43 |
| **RMSD from ideal values** | |
| **Bonds / Å** | 0.023 |
| **Angles / °** | 2.064 |
| **No. of TLS groups** | 2 |

**Table 5.2 Refinement statistics for pRET-KD:sunitinib**

Values for the outermost shell are shown in parenthesises.

### 5.3.3 pRET-KD:sunitinib structure analysis

The pRET-KD:sunitinib structure, as with all the RET-KD structures, displays a classical bilobal protein kinase structure (Figure 5.5). The smaller N-lobe consists of five anti-parallel beta strands (β1–5) and the catalytically important αC helix, between strands β3 and β4. The flexible glycine-rich loop (gly-rich loop), which lines the top of the active site, is found between β1 and β2.

In addition to the classic N-lobe features, RET-KD has an N-terminal helix dubbed αN, which contains 5 vector-derived residues (GPLSL) and 6 *de novo* RET residues (SVDAFK). Since the vector-derived residues are not similar to those found in RET (MENQV) it is unclear whether this helix is present in full length RET (discussed further in section 5.4.3.2). The linker (I711–P715) connecting the αN helix to the rest of the KD is disordered.

The linker between the N and C-lobes, referred to as the kinase 'hinge' since the relationship between the N and C-lobes is flexible, connects strand β5 to helix αD. The kinase hinge residues line the back of the nucleotide-binding pocket and engage ATP pocket inhibitors, including sunitinib.

The larger C-lobe of RET-KD is predominantly α helical, with 7 α helices (αD–αI) and contains the glycine and serine-rich kinase insert region, between αD and αE. Residues G823–Y826 and P841–E843 from the kinase insert, which are present in the protein construct, are absent from the refined electron density map and are therefore disordered in the crystal. The long linker region between αE and αEF, contains antiparallel beta strands β7 and β8, which form the base of the nucleotide binding pocket; and the activation loop, including the phosphorylated Y905, which is functionally required for RET kinase signalling *in vivo* (Kawamoto *et al.*, 2004). Y905 is the only visibly phosphorylated residue in this structure. Y900, the other invariant activation loop autophosphorylation site, is solvent exposed and the side chain is disordered.

Overall the pRET-KD:sunitinib structure presented here is very similar to the published phosphorylated and non-phosphorylated RET-KD structures. The root mean square deviation (RMSD) between pRET-KD:sunitinib and pRET-KD:AMP (PDB 2IVT) is only 0.206 Å over 264 Cα's. The RMSD between pRET-KD:sunitinib and non-phosphorylated RET-KD:PP1 (PDB 2IVS) is slightly greater at 0.410 Å over 260 Cα's. Furthermore the symmetric head-to-tail dimer seen in all published RET-KD structures is also present in the pRET-KD:sunitinib crystal, burying a surface area of 1382.6 $Å^2$. (The RET-KD crystallographic dimer is described in section 1.7.2).

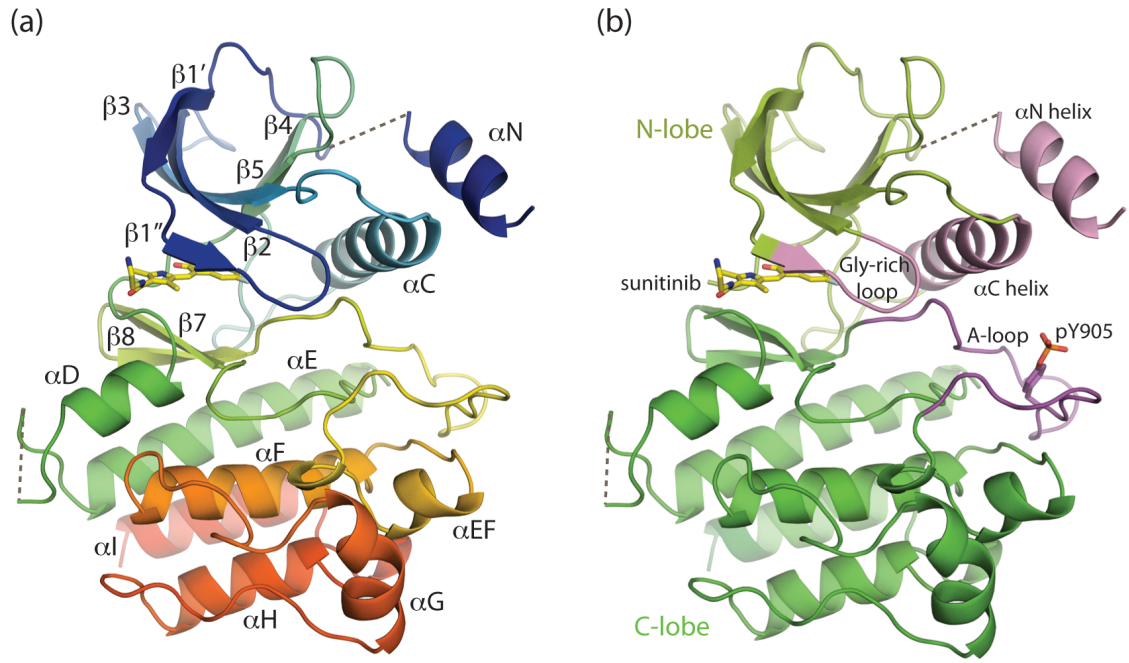



**Figure 5.5 pRET-KD:sunitinib structure key features**

The pRET-KD:sunitinib architecture is essentially the same as the published pRET-KD structures, with sunitinib bound in the ATP binding pocket. (a) Ribbon representation of the pRET-KD structure, with rainbow colouring from N to C termini. The α helices and β strands are annotated according to the standard kinase nomenclature. The two regions not seen in the electron density map, the αN linker and the kinase insert, are indicated by dotted lines. (b) The bilobate kinase structure is shown, with the N-lobe coloured light green (highlighted regions coloured light pink), and the C-lobe coloured green (highlighted regions in magenta). Key regulatory and RET specific elements are highlighted and labelled, with the crucial phosphorylated tyrosine (Y905) side chain shown in stick view. Sunitinib (coloured yellow) is also depicted in the stick view.

#### *5.3.3.1 pRET-KD bound to sunitinib adopts an 'active' kinase conformation*

The RET-KD molecule in the pRET-KD:sunitinib structure adopts an essentially catalytically competent, or 'active', conformation. Organisation of the kinase such that the catalytic site residues are properly positioned for catalysis involves the correct positioning of a number of structural elements within the KD. These are described below.

Firstly the αC helix is positioned close enough to the active site to allow E775 to hydrogen bond with K758, an interaction important for the orientating K758 for binding to the α and β phosphates of ATP. K758 side chain itself is in a slightly different

conformation to those observed in published ATP-analogue bound active RTK KD structures. This is likely due to the absence of ordered coordinating ATP-phosphates in pRET-KD:sunitinib.

The catalytic loop, located between αE and β7, contains the catalytic base D874, which is orientated by R878 and N879 to interact with the attacking hydroxyl from the substrate tyrosine side chain. In the pRET-KD:sunitinib structure this key residue is properly orientated when compared to equivalent residues in a substrate-peptide and ATP-analogue bound FGFR KD structure (PDB 2PVF).

In the pRET-KD:sunitinib structure the conserved DFG sequence motif at the start of the activation loop (following β8) adopts the *DFG-in* conformation consistent with an active kinase. The positioning of this motif is generally linked to the conformation of the whole activation loop in KD structures. The *DFG-in* conformation observed here positions D892 ready for magnesium coordination. The D892 side chain is rotated slightly relative to the adopted position in the presence of $Mg^{2+}$ and ATP seen in the FGFR structure, but the movement required is small.

The activation loop itself is ordered and displays catalytically competent conformation, which does not impinge upon substrate binding. As previously mentioned Y905 in the activation loop is phosphorylated. The phosphate group provides an anchor point for this conformation, making hydrogen bonds to R770, R897 and K907 side chains from the αC helix and within the activation loop. Although in RET-KD crystals Y905 phosphorylation is not required for adoption of this conformation (Knowles *et al.*, 2006).

All these features described for the pRET-KD:sunitinib structure are also observed in the published pRET-KD structures, excepting the positioning of K758, indicating that sunitinib is not effecting the overall arrangement of the pRET-KD structure. This shows that sunitinib is capable of binding to the active conformation of the RET-KD, thereby rendering the enzyme inactive solely by occupying the ATP-binding pocket. The details of sunitinib's binding to this pocket are described below.

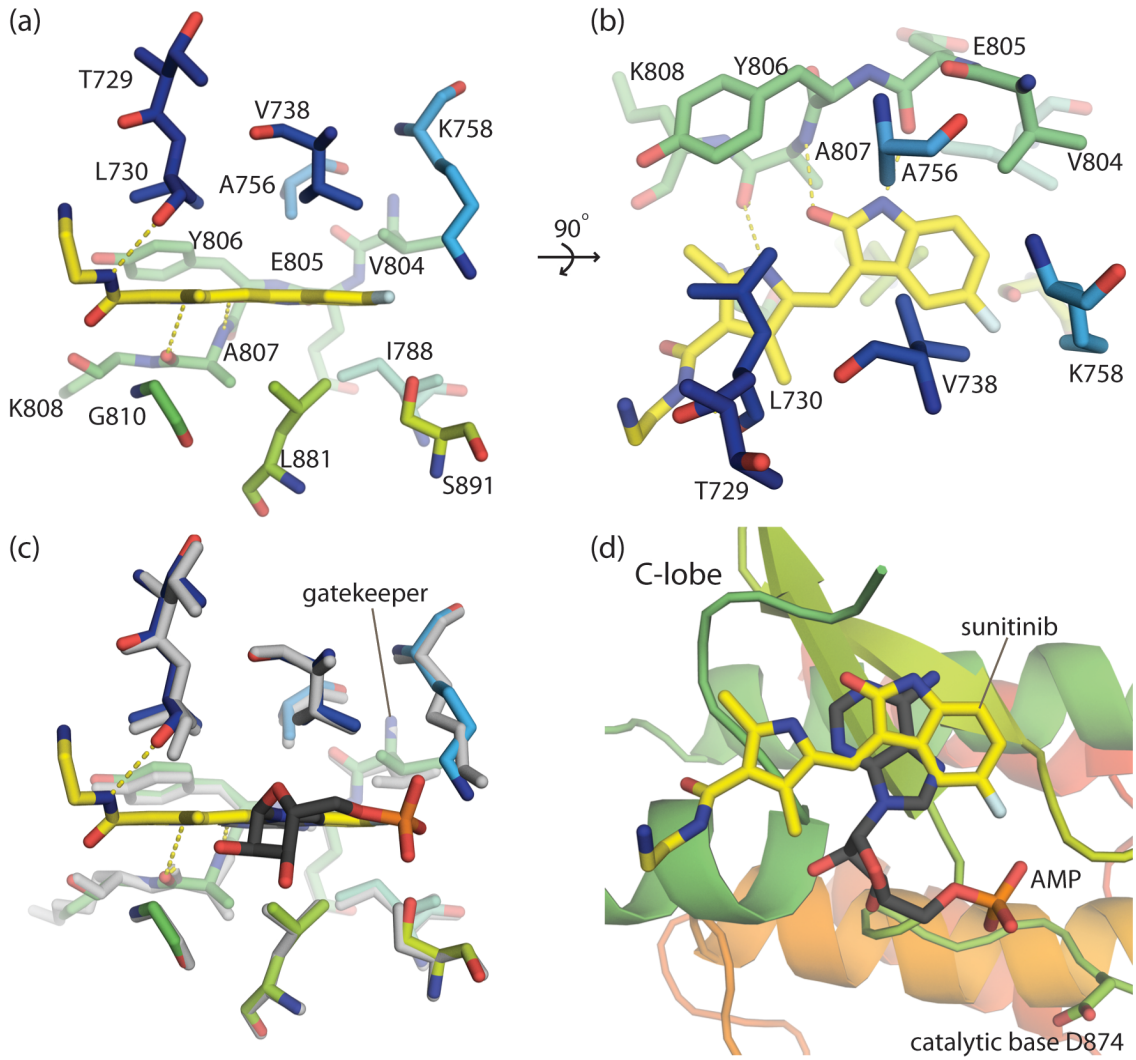



**Figure 5.6 Sunitinib-binding in the nucleotide pocket**

(a) & (b) Close up of the sunitinib-binding pocket. Sunitinib (yellow sticks) and residues within 4 Å of sunitinib in the crystal structure are shown. Predicted hydrogen bonds are depicted as dotted lines. (e) Overlay of sunitinib-binding pocket residues with those same residues from the AMP bound pRET-KD structure (grey sticks) and AMP (dark grey sticks). (d) Relative positions of AMP and sunitinib in pRET-KD crystal structures, viewed from above. The C-lobe of the pRET-KD is shown in ribbon representation.

#### *5.3.3.2 Sunitinib-binding in the ATP pocket*

Well-defined electron density was observed for sunitinib for all but two ethyl groups of the solvent exposed diethylamino moiety. Sunitinib binds RET in the hydrophobic cleft with the β1, β2 and β3 strands forming the roof (residues T729, L730, V738, A756 and K758); the kinase hinge at the back (V804–G810); and β7 and β8 strands at the base

(L881 and S891) (Figure 5.6a-d). Sunitinib forms four hydrogen bonds with RET. Three of these are between the sunitinib ring system and main chain groups from the kinase hinge: the indolylidine nitrogen to the E805 main chain carbonyl (2.71 Å N to O distance); the indolylidine group carboxyl to the Y806 main chain nitrogen (2.98 Å); and the pyrrole group nitrogen to the A807 main chain carbonyl (3.42 Å). The final hydrogen bond is between the nitrogen from the carboxyamide group on the sunitinib pyrrole ring and the N-lobe L770 main chain carbonyl (2.94 Å) (Figure 5.6c&d). These hydrogen bonds are all with main chain groups, which partly indicates the basis for sunitinib's lack of specificity for RET.

The planar ring system of sunitinib overlaps with the binding site for the adenine group of ATP (Figure 5.6e&f), but inserts further into the hydrophobic pocket pressing up to the V804 gatekeeper residue. Despite this, sunitinib-binding results in very little conformational change in the nucleotide pocket residues when compared with those from the AMP-bound structure (Figure 5.6e). A small movement of the V804 side chain to accommodate the indolylidine group, a slight movement of the main-chain carbonyl of L770 as a result of the hydrogen bond, and a different conformation of the flexible K758 residue due to the lack of an ATP phosphate group are the most noticeable changes.

### 5.3.4 Comparison with other sunitinib-bound KD structures

There are three sunitinib-bound protein kinase structures in the PDB: Two tyrosine kinases: the receptor tyrosine kinase KIT's KD (Gajiwala *et al.*, 2009), and the intracellular tyrosine kinase IL-2-inducible T cell kinase (ITK) (Kutach *et al.*, 2010); and one serine threonine kinase: phosphorylase kinase's catalytic subunit gamma 2 (unpublished).

The sunitinib-bound KIT-KD structure has the kinase activation loop in an inactive conformation, blocking the substrate-binding site, with the magnesium-coordinating DFG motif adopting a *DFG-out* conformation. Furthermore Gajiwala *et al.* show that sunitinib preferentially inhibits the non-phosphorylated and inactive (*DFG-out*) state of KIT-KD using an *in vitro* kinase assay. The observed IC50 values for non-

phosphorylated versus phosphorylated KIT-KD were 42 nM and 7 μM respectively (Gajiwala *et al.*, 2009).

As mentioned above, RET-KD exhibits an active *DFG-in* conformation in the sunitinib-bound structure (Figure 5.7a&b), as for all published RET-KD structures. To test whether there is any difference in sunitinib inhibition of kinase activity between non-phosphorylated and phosphorylated RET-KD, as seen for KIT, I performed a similar *in vitro* kinase assay. The poly$E_4$Y peptide was used as a substrate, and ATP consumption was measured at varying sunitinib concentrations (Figure 5.7c). The IC50 values for sunitinib with non-phosphorylated and phosphorylated RET were very similar: 0.56 ± 1.21 μM (IC50 ± standard error of the mean) and 1.44 ± 1.14 μM respectively. Therefore there does not seem to be an appreciable difference between sunitinib's inhibition of the two phosphorylation states of RET *in vitro*. This could reflect either a lack of specificity by sunitinib for a particular DFG conformation or the absence of a *DFG-out* conformation of RET-KD in solution. The difference in activity between phosphorylated and non-phosphorylated RET-KD by this assay is only 2.5–3×, and the non-phosphorylated RET-KD structure still shows an active activation loop structure with *DFG-in* (Knowles *et al.*, 2006). Therefore unlike KIT where the non-phosphorylated KD can be reasonably assumed to be predominantly in the *DFG-out* conformation, this cannot be assumed for non-phosphorylated RET-KD. As such the similarity between IC50 values for sunitinib with RET-KD in both phosphorylation states may reflect sunitinib-binding essentially the same conformation of the kinase—as observed in the crystal structures—with *DFG-in*.

Sunitinib targeting the active conformation of kinases is more consistent with it being a Type 1 multi-kinase ATP-competitive inhibitor. This is because the active conformation of the active site of protein kinases is very similar, whereas there is greater divergence between inactive conformations and mechanisms of autoinhibition (Huse and Kuriyan, 2002). Consistent with this the other two sunitinib-bound protein kinase structures exhibit an active activation loop conformation with *DFG-in* like the RET structure (Figure 5.7). Notably the tyrosine kinase ITK structure shows an active conformation despite the protein being non-phosphorylated. This is not due to a preference for the active conformation in the crystal, since in the same study two other inhibitors were also co-crystallised with ITK and both of these exhibited an inactive arrangement

(Kutach *et al.*, 2010). KIT may therefore be an exception to a general preference by sunitinib for the active conformation.

The major differences between the three sunitinib-bound active KD structures are seen in the positions of the flexible gly-rich loop and the αC helix (Figure 5.7d&e). Neither of these differences is likely to be linked to sunitinib-binding. The gly-rich loop of RET, as seen in all current RET-KD structures, collapses over the active site in such a way that would prevent ATP binding (were sunitinib not there). The gly-rich loop of ITK is disordered, and in phosphorylase kinase the gly-rich loop is raised away from the active site allowing ATP access. The αC helix contains magnesium-coordinating residues required for catalysis; therefore in active kinase structures the αC helix is positioned close the active site. The N-terminus of the RET αC helix is bent further towards the kinase cleft than in the other two *DFG-in* sunitinib-bound kinase structures, despite their active conformation (corroborated by the inactive KIT structure where the αC helix is even further from the catalytic centre). The αC helix in the active insulin receptor KD structure has the same arrangement as RET-KD however, so this is not a RET specific arrangement.

Comparison of the three kinases bound to sunitinib highlights the uncommon nature of the interaction of KIT-KD with sunitinib, and also highlights some of the macro structural differences between the RET-KD and other 'active' KD crystal structures.

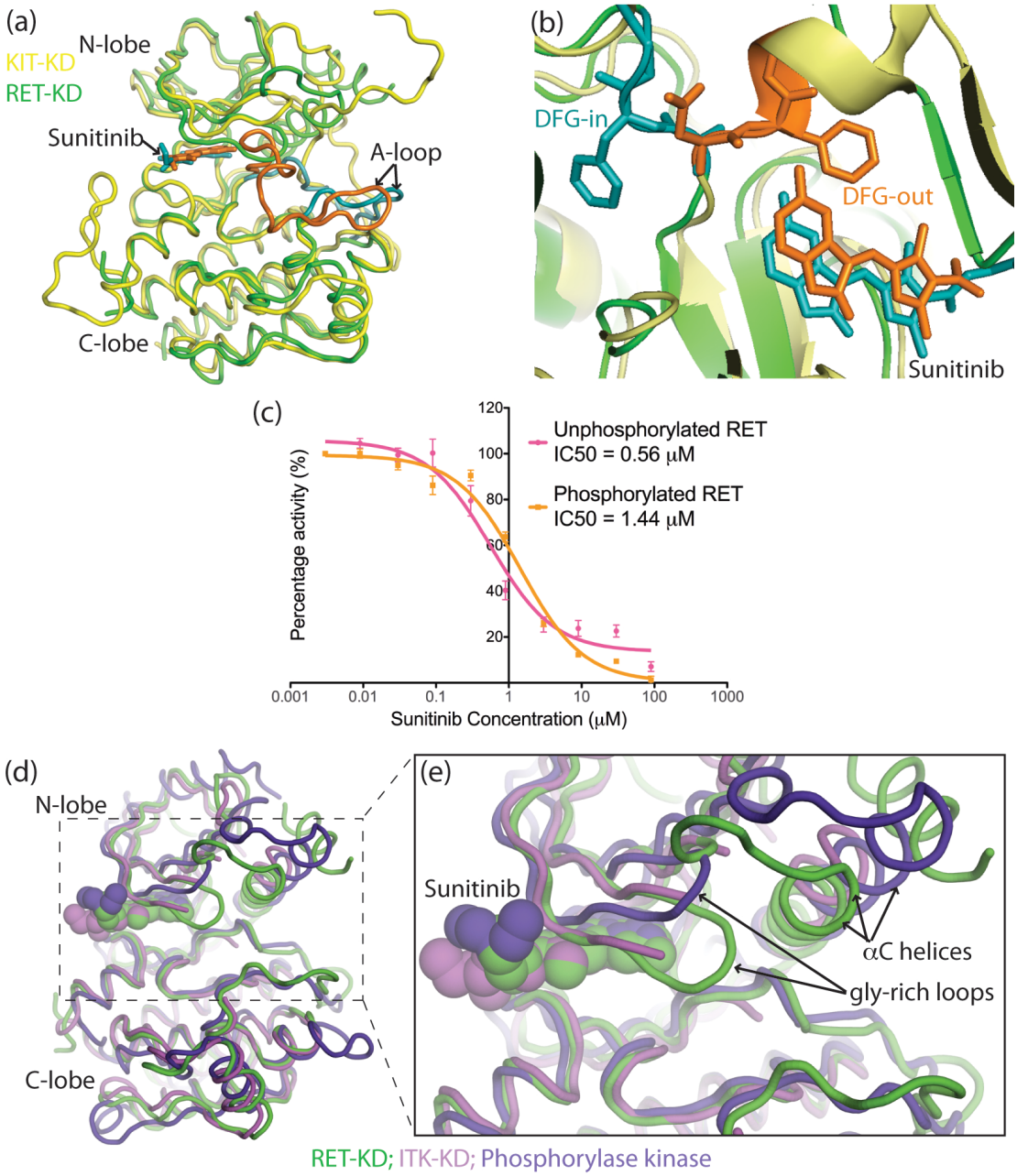



**Figure 5.7 Sunitinib can bind both active and inactive kinases**

(a) Superposition of sunitinib-bound RET-KD (green ribbon and highlighted with teal) and KIT-KD (PDB 3G0E, yellow ribbon and highlighted with orange) revealing the different activation loop (A-loop) conformations. (b) Close up, from above, of sunitinib binding RET-KD in the active *DFG-in* conformation and KIT-KD the inactive *DFG-out*. Sunitinib, and the DFG motifs at the start of the activation loop, are highlighted and shown as sticks. (c) Results of an *in vitro* ADP-coupled kinase assay comparing sunitinib inhibition of phosphorylated and unphosphorylated RET-KD kinase activity towards a poly$E_4Y$ peptide substrate. (d) & (e) Superposition of the sunitinib-bound 'active' KD structures, depicted as ribbons with sunitinib shown as spheres. RET-KD is coloured green, IL-2-inducible T cell kinase (ITK), magenta (PDB 3MIY), and phosphorylase kinase gamma 2, purple (PDB 2Y7J).

## 5.4 A high resolution adenosine-bound RET-KD structure

### 5.4.1 Rationale and crystallisation

The RET-KD has been co-crystallised with a number of inhibitors, as discussed in the previous section, along with two ATP derivatives: AMP and cyclic AMP (Knowles *et al.*, 2006). No structures with an ATP-analogue bound have so far been published, and in all RET-KD structures the gly-rich loop is in a 'closed' conformation (collapsed over the active site) precluding ATP-binding.

A structure containing all three phosphate groups of ATP would be desirable, since this would best represent the enzyme in a catalytically productive state, and may help trap a substrate-peptide engaged by the kinase. I therefore conducted crystal trials using RET-KD mutations discussed in section 5.5 with AMPPNP. These mutants were produced in an attempt to obtain a different crystal form, and since it was unknown whether the current crystal form—obtained in the sodium formate condition—would accommodate an ATP-analogue these mutants were a possible way of obtaining an ATP-analogue bound structure.

As a control I set up the wild type pRET-KD with AMPPNP in the sodium formate condition, to see whether RET-KD could crystallise in this condition when bound to AMPPNP. Sitting drops of 3 mg/ml pRET-KD with 3 molar excess AMPPNP were set up against a well solution of 0.1 M sodium citrate pH 5.5, and 1.8–2.3 M sodium formate. No crystallisation was observed after 6 days, so I tried seeding some of the drops using crushed pRET-KD:sunitinib crystals. This did result in the formation of plate crystals, as seen for pRET-KD:sunitinib, but only in the seeded drops. In house testing for X-ray diffraction gave spots to around 4 Å.

Due to other priorities, these crystals were not pursued further at the time. However upon looking at the tray again 20 months later, crystals had formed in the unseeded drops. In particular, one large 3D single crystal had formed in the 2.1 M sodium formate condition (Figure 5.8). This crystal was harvested, cryoprotected with mother liquor plus 25% glycerol and flash frozen in liquid nitrogen.

### 5.4.2 Structure determination of pRET-KD:adenosine

#### *5.4.2.1 Data Collection*

This single crystal was taken to Diamond Light Source and Dr. Andrew Purkiss-Trew collected diffraction data. Test shots showed diffraction to around 1.6 Å, although only data to 1.8 Å was used due to low completeness (Figure 5.8 and Table 5.3). A dataset was collected using the Pilatus detector over 102°, with an oscillation of 0.2° per image. The data showed some low-level ice contamination, particularly in one crystal orientation. This is the highest resolution dataset obtained for RET-KD and we were keen to see if this dataset would allow identification of flexible regions, disordered in previous structures.

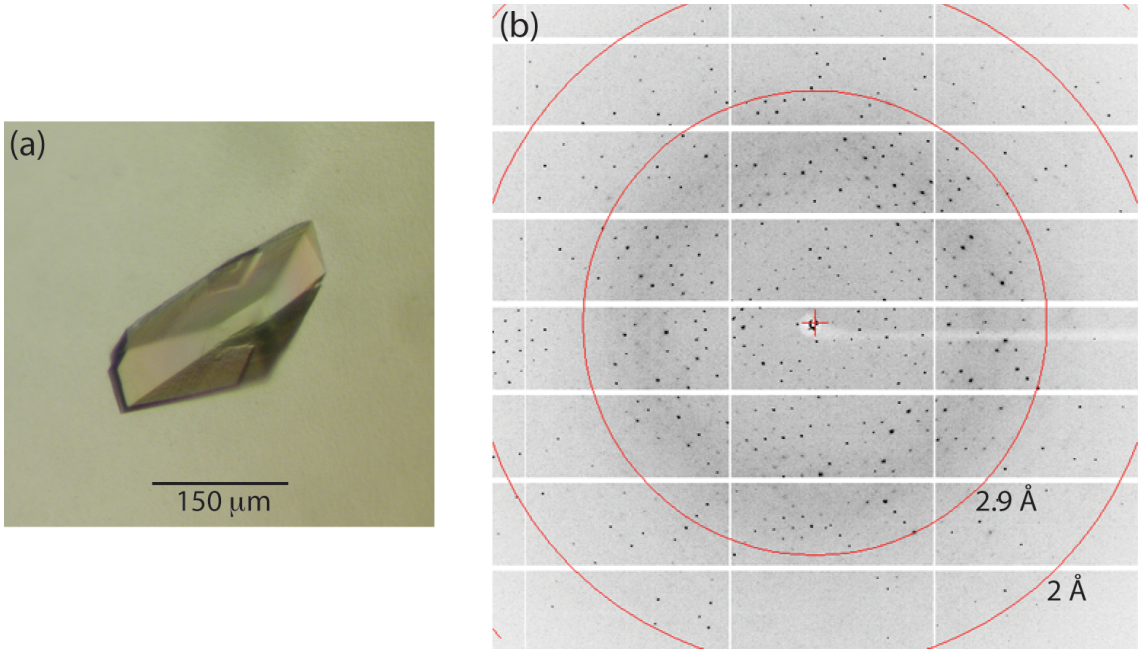



**Figure 5.8 pRET-KD:adenosine crystal and X-ray diffraction pattern**

(a) Image of 300 µm pRET-KD:adenosine crystal grown in 2.1 M sodium formate, 0.1 M sodium citrate pH 5.5, 3mg/ml protein. (b) Representative diffraction pattern image, collected on a Pilatus detector at Diamond Light Source I-24.

| Data collection | |
|---|---|
| **X-ray source** | Diamond Light Source I-24 |
| **Wavelength / Å** | 0.9700 |
| **Space group** | C2 |
| **Unit cell dimensions** | |
| **a, b, c / Å** | 73.162, 68.980, 78.631 |
| **α, β, γ / °** | 90.00, 101.78, 90.00 |
| **Resolution / Å** | 39.24–1.80 (1.86–1.80) |
| **Total reflections** | 64144 |
| **Unique reflections** | 33908 |
| **Mosaicity / °** | 0.57 |
| **Redundancy** | 1.89 (1.93) |
| **Completeness / %** | 95.4 (94.2) |
| **I/σ** | 13.6 (1.8) |
| **$R_{merge}$ / %** | 3.7 (29.2) |

**Table 5.3 Data collection statistics for pRET-KD:adenosine**

Values for the outermost shell are shown in parenthesises.

#### *5.4.2.2 Solving the structure by molecular replacement*

D*TREK (Pflugrath, 1999) was used to process the data, since it deals with Pilatus data particularly well. The spacegroup was determined to be C2, with cell parameters as detailed in Table 5.3. After removal of a single poor image, and cutting back the resolution to 1.6 Å to give an I/σ of 1 and $R_{merge}$ of 47% at the highest resolution shell, the data was integrated, scaled and merged in D*TREK.

Molecular replacement was carried out using the CCP4 suite (Collaborative Computational Project, 1994). Cell content analysis predicted that there was one molecule in the asymmetric unit with a 60% solvent content. The RET-KD protein coordinates from the pRET-KD:PP1 structure (PDB 2IVV) were used as the search model for molecular replacement in PHASER (McCoy *et al.*, 2007). A single good solution was found.

#### *5.4.2.3 Refinement using PHENIX and BUSTER*

The structure was refined initially using the maximum-likelihood automated refinement program PHENIX.REFINE (Adams *et al.*, 2010) with model building carried out in COOT (Emsley and Cowtan, 2004). Refinement was carried out using three TLS groups: residues 700–713 (αN helix), 715–805 (N-lobe), and 811–1012 (C-lobe). The

improvement in the $R_{factors}$ stalled during refinement. We reasoned this could be due to poor data quality at high resolution and the data was therefore cut back to 1.8 Å, which resulted in a moderate improvement in $R_{work}$ and $R_{free}$. Electron density for adenosine in the nucleotide pocket was observed and built, but throughout refinement no density for the phosphate groups of AMPPNP was seen. Electron density for 10 formate molecules and 2 glycerols was also observed and these were added. Previous structures only contained between two and four formate molecules; the number here is likely a reflection of the higher resolution of this structure.

The gly-rich loop was found to be in two different conformations, and the electron density for this region was too poor to build in both conformations with confidence. A different refinement program, BUSTER (Blanc *et al.*, 2004), was therefore tried. BUSTER uses maximum-likelihood and maximum entropy techniques for refinement. It was designed to overcome the bias towards the modelled incomplete structure, which occurs in other refinement programs, and limits the improvement in the electron density map for unmodelled regions of the structure. We therefore tried BUSTER to see if it would improve the electron density for the gly-rich loop. Refining the structure with BUSTER did improve map connectivity, and particularly the gly-rich loop density, although the overall $R_{factors}$ were not improved. Six rounds of model building and refinement in BUSTER (with TLS groups as before) were carried out, resulting in a final structure with $R_{work}$ = 19.8% and $R_{free}$ = 21.3% (Table 5.4).

| Refinement | |
|---|---|
| **Resolution / Å** | 39.28–1.80 (1.85–1.80) |
| **Completeness** | 95.27 (95.27) |
| **No. of reflections** | 33858 (2898) |
| **$R_{work}$ / %** | 19.77 (24.23) |
| **$R_{free}$ / %** | 21.33 (28.66) |
| **Correlation coefficient (Fo-Fc) / %** | 95.68 |
| **Total protein atoms** | 2317 |
| **Total ligand atoms** | 182 |
| **Total solvent atoms** | 105 |
| **B-factor (Å²)** | 43.28 |
| **Protein** | 42.42 |
| **Ligand** | 44.40 |
| **Water** | 52.56 |
| **Ramachandran Plot / %** | |
| **Favoured** | 97.38 |
| **Allowed** | 2.25 |
| **Outliers** | 0.37 |
| **RMSD from ideal values** | |
| **Bonds / Å** | 0.01 |
| **Angles / °** | 0.97 |
| **No. of TLS groups** | 3 |

**Table 5.4 Refinement statistics for pRET-KD:adenosine**

Values for the outermost shell are shown in parenthesises.

### 5.4.3 Analysis of the pRET-KD:adenosine high resolution structure

The refined 1.8 Å structure from the 20 month old crystal grown from pRET-KD + AMPPNP, showed that only adenosine was present in the nucleotide pocket. Presumably the phosphodiester bonds had undergone water-mediated hydrolysis over time leaving just adenosine.

The overall structure of the kinase is very similar to the sunitinib-bound structure presented in the previous section, and the AMP bound pRET-KD structure (PDB 2IVT), which only differs in the presence of a single phosphate on the ligand. The RMSD is 0.267 Å over 248 Cα's between the sunitinib and adenosine structures, and 0.231 Å over 243 Cα's between the AMP and adenosine structures. As with these structures RET contains a single phosphorylated tyrosine residue on the activation loop, pY905. As expected the activation loop is in an open, ordered, 'active' conformation (Figure 5.9) with *DFG-in*. Unsurprisingly the adenosine-bound pRET-KD crystals contain the

same symmetric head-to-tail dimer seen in the pRET-KD:sunitinib and other RET-KD structures.

The questions of whether an ATP-analogue can be accommodated in this crystal lattice and what conformational changes in the gly-rich loop would be required for ATP accommodation are not answered by this structure. Fresh pRET-KD:AMPPNP crystals with better diffraction than those produced in this study (section 5.4.1) are needed. However the pRET-KD:adenosine structure does contain some notable differences and additions, due to the improved resolution and original presence of AMPPNP, compared to previously solved RET-KD structures. These include electron density for the linker between the αN helix and the rest of the KD, and two distinct conformations of the gly-rich loop (Figure 5.9). These features are discussed further below.

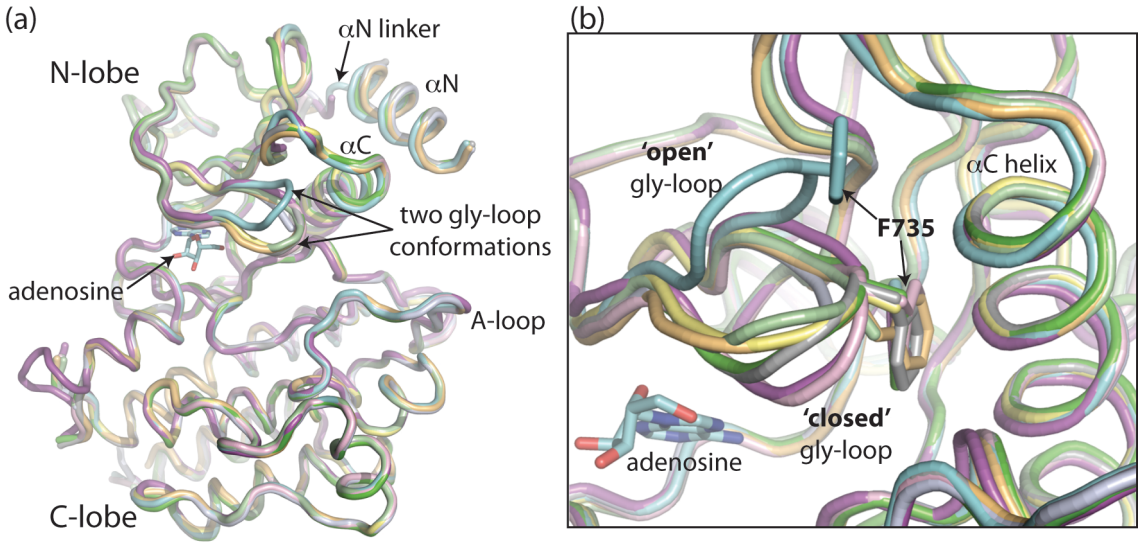



**Figure 5.9 Comparison of all published RET-KD structures showing the unique features in the adenosine-bound structure**

(a) Alignment of all published RET-KD structures along with the two presented in this thesis. Protein structures shown as tubes: pRET-KD:adenosine, cyan; pRET-KD:sunitinib, green; non-phosphorylated RET-KD:cAMP (2IVS), magenta; pRET-KD:AMP (2IVT), grey; pRET-KD:PP1 (2IVV), pale yellow; pRET-KD:ZD6474 (2IVU), pale green; pRET-KD:X2K (2X2K), pale pink; pRET-KD:X2L (2X2L), light orange; pRET-KD:X2M (2X2M), pale blue. Adenosine is shown as cyan sticks to mark the nucleotide pocket. (b) Close up of the 'open' and 'closed' gly-rich loop conformations, with the F735 side chain shown in stick depiction.

#### *5.4.3.1 Multiple conformations of the flexible gly-rich loop*

The gly-rich loop is known to be flexible and has been seen in a variety of positions in KD crystal structures, ranging from an 'open' conformer positioned away from the kinase cleft to a 'closed' conformer with the gly-rich loop folded down over the cleft, sterically blocking binding of ATP phosphate groups. In active ATP-bound KDs the gly-rich loop engages the nucleotide's phosphates and facilitates proper positioning of the γ-phosphate for phosphoryl transfer. An 'open' conformer of the gly-rich loop is proposed to facilitate ADP-release (Khavrutskii *et al.*, 2009). ATP-pocket inhibitors can also influence the conformation of the gly-rich loop, with the conserved phenylalanine (F735 in RET) often folding back to contact hydrophobic moieties in the bound inhibitor (Johnson, 2009). However in all the published RET-KD structures the gly-rich loop is in the same 'closed' position, with some slight variation evident (Figure 5.9b). This loop is present in two conformations in the pRET-KD:adenosine structure. One of which is similar to that seen in other RET-KD structures, with F735 packed against the αC helix, but the alternate is a new 'open' conformation, with F735 accessible. This 'open' conformation has the gly-rich loop bent away from the kinase cleft and the active site.

As is shown in Figure 5.10, overlaying the pRET-KD:adenosine structure with an ATP-analogue and substrate-peptide bound KD structure of the closely related FGFR2 (Chen *et al.*, 2007), reveals that the usual 'closed' conformation of the RET gly-rich loop in the crystal, as exemplified by the pRET-KD:AMP structure, would clash with the β and γ phosphates from ATP (Figure 5.10b). The usual gly-rich loop conformation is therefore an inhibitory conformation since it would impede ATP (or ADP) binding.

The alternative 'open' conformation of the gly-rich loop in the pRET-KD:adenosine structure is similar to the conformation of this loop in the pFGFR2:ATP-analogue:substrate peptide structure, showing that this conformation could accommodate ATP, and could therefore be described as an ATP-competent conformation. It does not precisely match the conformation seen in the FGFR2 structure however, due to the lack of the hydrogen bonds between the γ phosphate oxygens and gly-rich loop main chain nitrogens. Instead this 'open' conformation of the gly-rich loop is stabilised by two hydrogen bonds with the loop that precedes the αC helix,

located above the gly-rich loop. These H-bonds are formed between the main chain carbonyls of the gly-rich loop and the side chain nitrogen of N763 and the main chain nitrogen of K761.

The two different gly-rich loop conformations of RET are defined by the conformation of two residues (Figure 5.10c). One is E768 from the αC helix, which points upward when the gly-rich loop is in the more commonly observed 'closed' position, a conformation stabilised by a hydrogen bond between one of its side chain oxygens and the main chain nitrogen from S765 at the top of the αC helix. However this conformation is not compatible with the gly-rich loop in the 'open' position, with the side chain clashing with the gly-rich loop F735 side chain. The E768 side chain therefore adopts a different conformation when the gly-rich loop is in the 'open' position. In this case the side chain is pointing down in a way that would be incompatible with the 'closed' position of the gly-rich loop. The E768 rotamer is therefore linked to the position of the gly-rich loop. Similarly the F735 side chain must adopt a different rotamer in each of the two gly-rich loop conformations to avoid a clash with the αC helix.

The required conformations of E768 and F735 to accommodate the gly-rich loop 'open' position in RET are not required for the equivalent residues in FGFR2. This is due largely to the different position of the top end of the αC helix. The αC helices of RET and FGFR2 overlay well towards the base of the helix, where the magnesium coordinating glutamate resides (E775 in RET), however the top part of the helix in FGFR2 is bent away from the gly-rich loop. The gly-rich loop of FGFR2 is also slightly lower down and further away from the αC helix than RET. This means that although the D527 side chain (equivalent to RET E768) is pointing upwards it is not clashing with its gly-rich loop in the 'open' position. The side chain of F492 (equivalent to RET F735) is disordered and there is space for it to adopt either rotamer observed in RET.

Intriguingly E768D is an oncogenic mutation, although how it activates RET is unknown, particularly since it is such a subtle and conservative mutation. Its location close to the hypothesised inhibitory dimer interface, led to speculation that the presence of the shorter aspartate could weaken this interaction (Knowles *et al.*, 2006), but the glutamate side chain doesn't directly contact the other molecule in the dimer. If the

switching of E768 side chain rotamer is necessary for the gly-rich loop to adopt the active ATP accessible ‘open’ conformation, perhaps the smaller aspartate side chain more readily allows this enzyme to assume this conformation.

The reason both conformations are present in this pRET-KD:adenosine structure, is likely due to there originally being AMPPNP bound, which would require the gly-rich loop to be in the ‘open’ position. Over time water-mediated hydrolysis of phosphodiester bonds led to loss of all three phosphates, allowing the gly-rich loop to flip over into the kinase cleft. The steric constraint imposed by the necessity of the rotamer change of the F735 and E768 to allow this movement to occur in the crystal, may be the reason why the gly-rich loop is still found in the ‘open’ position at least 50% of the time.

In conclusion the pRET-KD:adenosine structure presented here reveals two conformations of the gly-rich loop. The first, seen in all other RET-KD structures, with the loop bent into the kinase cleft is incompatible with ATP binding. Whereas the second with the loop bent upwards, would be compatible with ATP binding, and therefore reveals a hitherto unseen active arrangement. The movement of the gly-rich loop is constrained by E768, an oncogenic mutation site, and therefore the E768D mutation may increase kinase activity by allowing the gly-rich loop to switch conformation more readily.

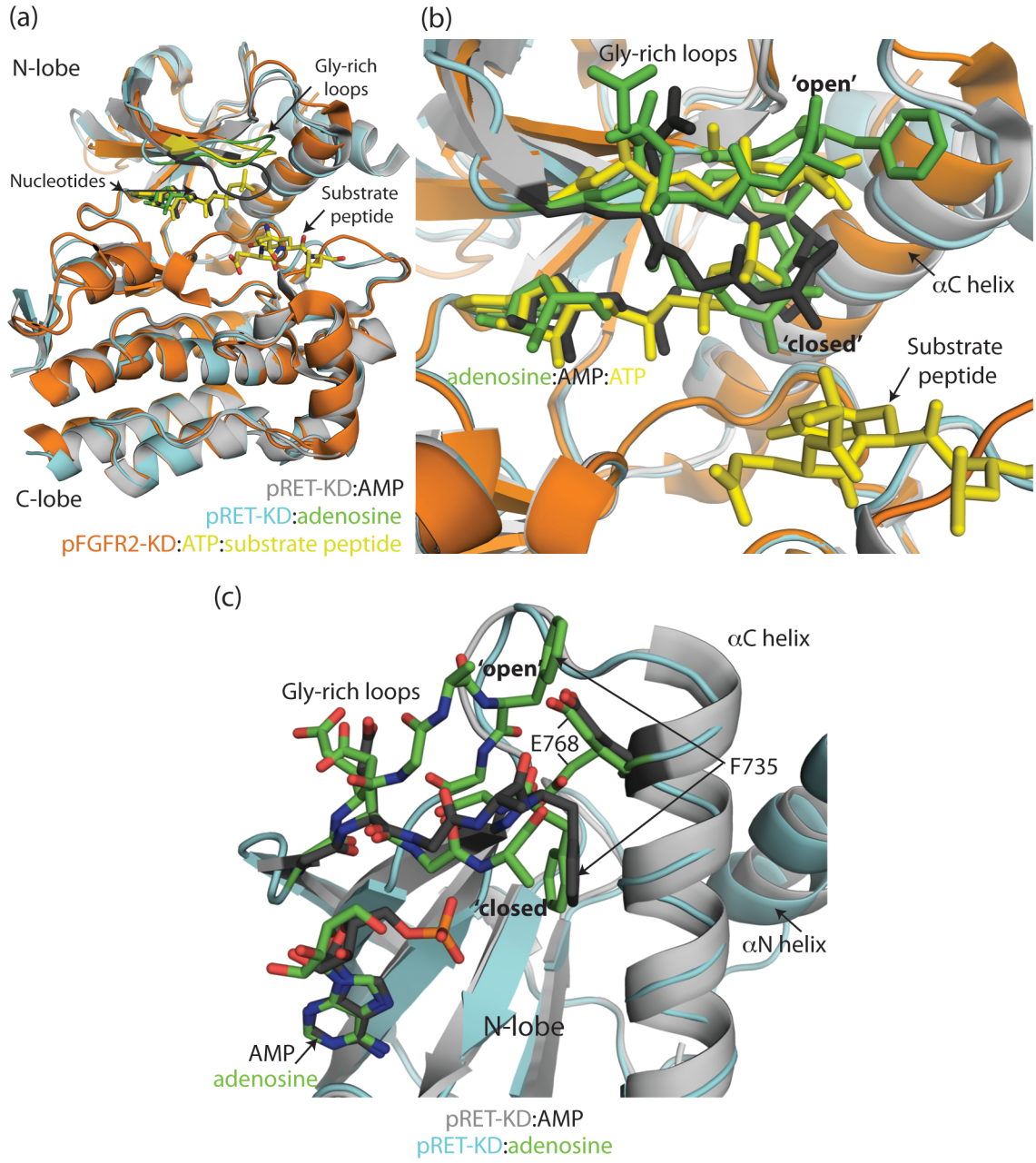



**Figure 5.10 'Open' and 'closed' conformations of the gly-rich loop**

The adenosine RET-KD structure contained two different gly-rich loop conformations. (a) Superposition of pRET-KD:adenosine structure (coloured cyan, highlighted features in green), AMP-bound pRET-KD structure (PDB 2IVT, coloured light grey, highlighted features in dark grey) and ATP-analogue and substrate-peptide bound pFGFR2-KD structure (PDB 2PSQ, coloured orange, highlighted features in yellow). (b) Close-up of the gly-rich loops from these structures, showing the inability of RET to accommodate ATP with the gly-rich loop in the 'closed' position. (c) View of the gly-rich loops of the two RET structures from beneath the αC helix. The two positions of the gly-rich loop in the adenosine-bound structure are reflected in the rotamer switch of F735 and the two conformations of E768 (side chain shown) from the αC helix.

#### *5.4.3.2 First complete look at the N-terminal linker preceding the first beta-strand of the N-lobe*

The pRET-KD:adenosine structure contains 10 ordered formates and 2 glycerols. One of the glycerols is bound in a pocket beneath the linker connecting the αN helix to the rest of the KD (αN linker). The entire linker is uniquely present in this structure (Figure 5.9), though with relatively high B factors for the previously absent residues (I711–P715 Cα's 70–78 $Å^2$). The bound glycerol molecule from the cryoprotectant is likely contributing to loop stabilisation (Figure 5.11).

The αN helix itself provides an anchor point for this linker. This helix packs against a hydrophobic surface formed the entire length of the exposed face of the αC helix. The αN helix residues L702, V706 and F710, lie close to the hydrophobic side chains of αC helix residues P766, L769, L773 and F776, and potentially L800 from the β4–β5 linker, stabilising both helices. However as previously mentioned (section 5.3.3) the first five residues of the αN helix are pBacPAK vector-derived, and unrelated to the equivalent residues in RET. These residues form the first turn of the αN helix, and include L702. The equivalent RET residue is an asparagine, which wouldn't fully recapitulate the hydrophobic interactions of L702. The presence of these vector-derived residues means that it is uncertain whether the αN helix is a true feature of the RET structure or is an artefact. An unpublished RET-ICD structure, including the native JM sequence, by Dr. Ivan Plaza-Menacho, shows an extended conformation of this region of RET that also presents hydrophobic contacts to the αC helix. The αN helix is therefore likely highlighting a true protein-protein interaction surface. Furthermore engagement of the hydrophobic surface of the αC helix is a regulatory mechanism in other kinases, including the RTK EGFR (Jura *et al.*, 2011), and could be regulatory in RET.

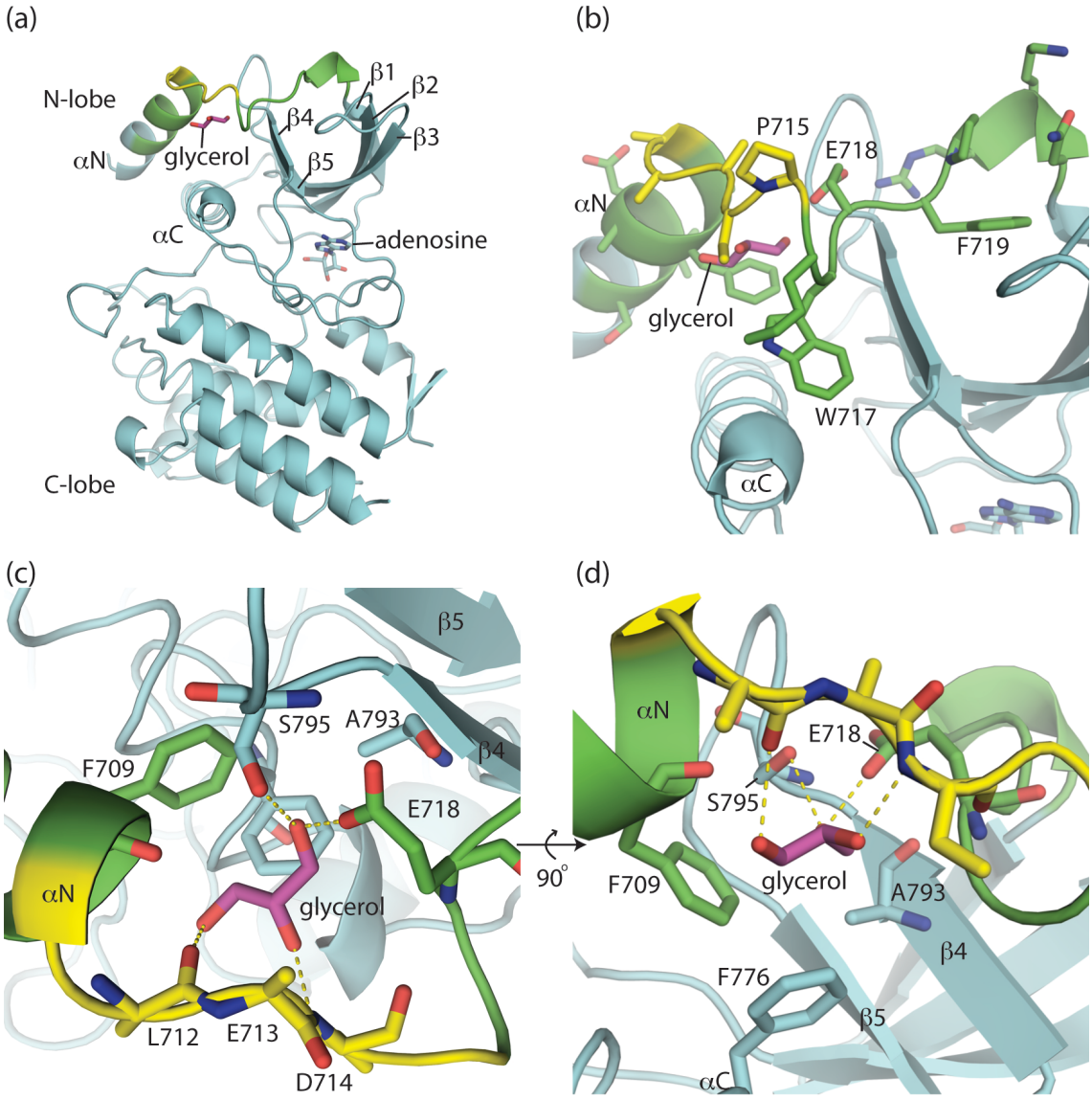



**Figure 5.11 RET-KD N-terminal linker**

(a) Ribbon representation of the pRET-KD:adenosine structure (cyan), with N-lobe structural elements labelled. The RET-derived N-terminal linker residues, preceding the N-lobe β1 strand, are coloured green, with the previously unseen residues in yellow. The glycerol molecule bound in the pocket formed by this linker, the β4 and β5 strands, and the αC helix is shown in stick view, in pink. (b) Close up of the N-terminal linker region, with the side chains shown as sticks. (c) & (d) The glycerol binding pocket, with residues within 4 Å of the glycerol shown as sticks. The hydrogen bonds are shown as yellow dotted lines.

A central anchor point for the αN linker is W717, which packs against the start of β4 near the base of the αC helix (Figure 5.11). This residue, which is conserved among FGFR kinases as well as RET, caps the regulatory spine in these kinases. This spine of

hydrophobic residues, formed when the DFG motif is in the catalytically competent *DFG-in* conformation, is conserved throughout protein kinases, and is proposed to be important for stabilising the active conformation of the kinase, since it connects the αC helix of the N-lobe, via the activation loop DFG motif F893, to the C-lobe αF helix (Kornev *et al.*, 2008) (Figure 5.12a). The presence of W717 at the top may confer additional stability to the spine and may also provide an additional source of regulation if its position can be altered. The spine residue it packs against, L790, is an oncogenic mutation site, and links W717 to L779 in the αC helix. The oncogenic mutation is L790F, and so perhaps the phenylalanine side leads to increased kinase activation by conferring greater stability to the spine than the native leucine.

The section of the αN linker between the αN helix and W717 has not previously been seen in its entirety, and even in this structure there is electron density only for the main chain of I711–D714, and the exact position of the P715 ring was difficult to model. The glycerol (shown in Figure 5.11c&d) ordering this linker is bound in a pocket with the αC helix at the base, the β4–β5 loop to one side, the αN helix to another, and the αN linker covering another side and the top. Glycerol forms four hydrogen bonds with pocket residues: Two with the main chain and one with the E718 side chain from the αN linker and the fourth with the S795 side chain from the β4–β5 linker. The base of this pocket is hydrophobic with F709, F776 and A793 all within 4 Å of the glycerol.

Further along the linker F719, also conserved with FGFR, provides an anchor point, for the highly charged E718–N723 section (Figure 5.11b). This residue packs into another hydrophobic pocket formed by residues from β1, β2, β5 and the β2–β3 linker (L724, A743, I803 and L746).

In conclusion the αN linker and the αN helix hide otherwise exposed hydrophobic surfaces from the RET N-lobe, with W717 and the αN helix hydrophobic residues (or the extended juxtamembrane domain hydrophobic residues) potentially playing a role in regulation of catalytic activity by their interactions with the regulatory spine and the αC helix respectively.

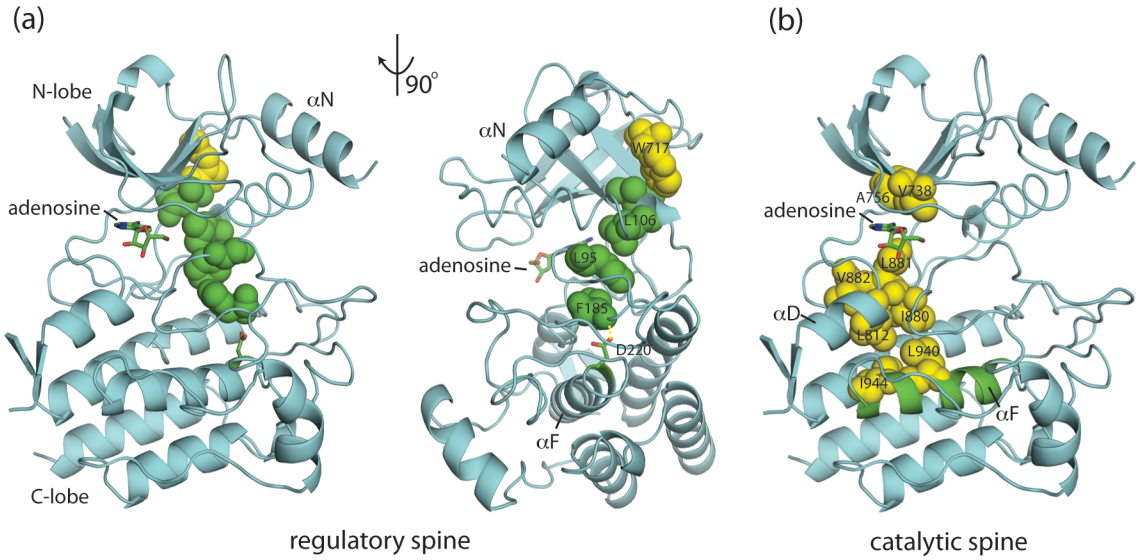



**Figure 5.12 Hydrophobic regulatory and catalytic spines**

Ribbon representation of the adenosine (green sticks) bound pRET-KD structure (cyan), with the hydrophobic residues forming the regulatory (a) and catalytic (b) spines shown as green and yellow spheres respectively. (a) RET (and its close ortholog, FGFR) contain a conserved tryptophan (W717, yellow spheres) found in the N-terminal linker, which lies at the top of the regulatory spine. (b) The adenine ring of adenosine completes the catalytic spine of the RET-KD, anchored by the αF helix (green).

## 5.5 Accessing a different RET-KD crystal form

All the RET-KD structures, including the two presented here, display the same head-to-tail dimer in the crystal. Along with this there are very few differences between the structures, aside from the gly-rich loop conformation discussed above. In particular the activation loop is in an active conformation in all the structures with *DFG-in*. This is frustrating since seeing the kinase in different conformations allows much greater insight into the regulatory mechanisms that govern its activity. Furthermore, although the active conformation of proteins kinases across the superfamily is very similar, there is much greater divergence in the structure of their inactive conformations, which are therefore a promising target for designing specific inhibitors for use in the clinic. Obtaining the structure of a kinase in its inactive form can both reveal the kinase specific mechanisms involved in its regulation and provide a basis for rational drug design.

It may be a constraint of the crystal lattice that allows only the active state of the kinase to be observed in this crystal form, which is corroborated by our inability to produce

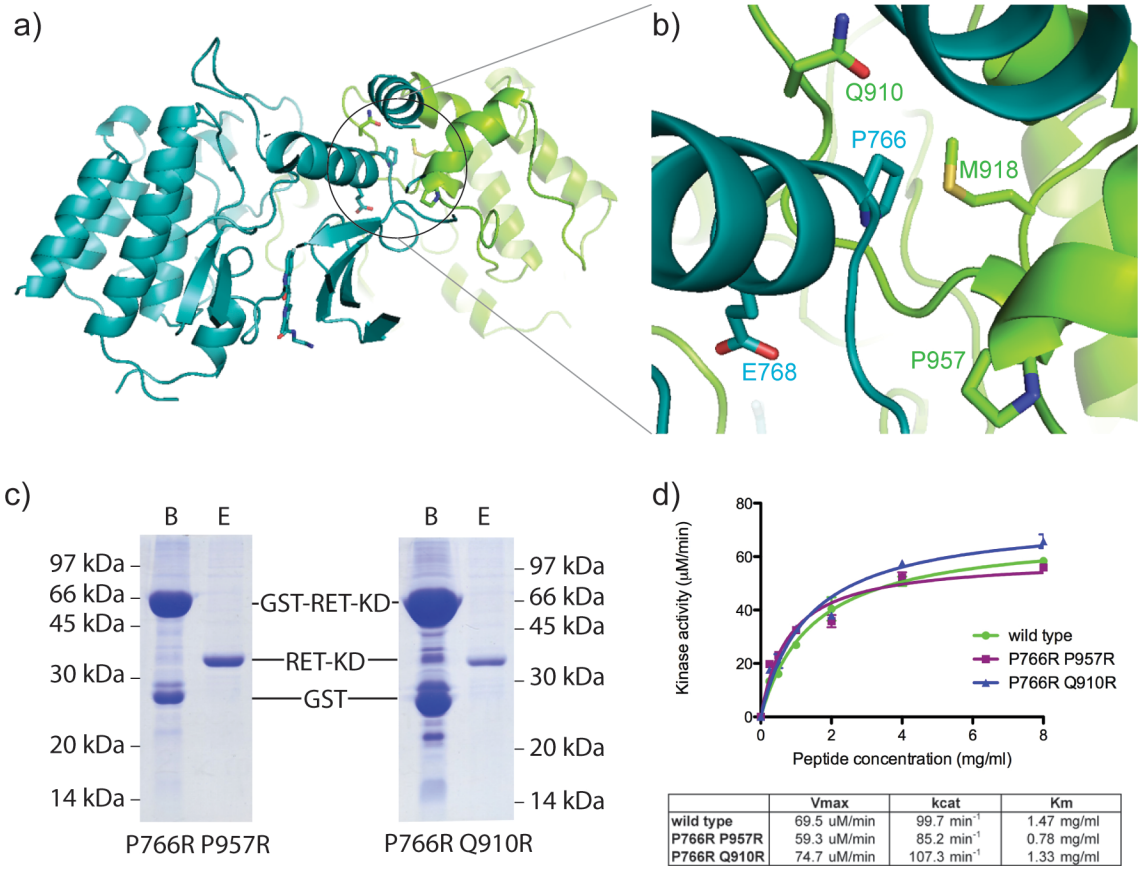
crystals with known *DFG-out* targeting inhibitors such as sorafenib. Since the dominant interaction in the crystal lattice is the head-to-tail dimer interface, we hypothesised that generating mutants, which weakened this interaction, may improve the chances of obtaining a crystal form with a different arrangement of the kinase. A further motivation for producing mutants that weakened the dimer interaction was to generate tools for testing the '*trans*-inhibitory' dimer hypothesis (discussed in section 1.7.2) in cell-based assays.

### 5.5.1 Design and production of 'dimer-deficient' mutations

Crystal structures of the MEN2B mutation, M918T, and the harsher P766R mutation showed these mutations failed to disrupt dimer formation in the crystal, although the P766R mutation did result in a movement of the two monomers in the dimer relative to each other (unpublished data). I therefore identified two additional residues—Q910 and P957—at the dimer interface, mutation of which to arginines in a P766R context was hoped would result in complete disruption of the dimer in crystals (Figure 5.13a&b). I therefore produced two double mutant RET-KD constructs RET-KD$_{P766R/Q910R}$ and RET-KD$_{P766R/P957R}$, dubbed dimer-deficient mutants.

Baculoviruses expressing the GST-RET-KD dimer-deficient mutants were produced by the protein production facility. Large-scale expression trials in Hi5 cells showed similar levels of expression to the wild type protein (up to 4 mg/l), and affinity purification resulted in stable, pure protein (Figure 5.13c).

Having produced the dimer-deficient mutants I first tested whether kinase activity was intact. These mutations were designed to disrupt dimer formation, which may be relevant to kinase activity *in vivo*, however RET-KD is a monomer in solution and therefore these mutations should not impact kinase activity in solution. Kinase activity was tested in an *in vitro* ADP-coupled kinase assay using a peptide substrate (polyE$_4$Y). The activity of the phosphorylated RET-KD dimer-deficient mutants was indistinguishable from phosphorylated wild type RET-KD showing that the kinase was still fully functional in solution (Figure 5.13d).




| | Vmax | kcat | Km |
|---|---|---|---|
| wild type | 69.5 uM/min | 99.7 $min^{-1}$ | 1.47 mg/ml |
| P766R P957R | 59.3 uM/min | 85.2 $min^{-1}$ | 0.78 mg/ml |
| P766R Q910R | 74.7 uM/min | 107.3 $min^{-1}$ | 1.33 mg/ml |

**Figure 5.13 Design and production of 'dimer-deficient' mutants**

(a) Head-to-tail dimer of RET-KD seen in all current crystal structures. The dimer from the pRET-KD:sunitinib structure shown in the ribbon representation with one monomer in green and the other teal. (b) Close up of the dimer interface between the N-lobe of one monomer (teal) and the C-lobe of the second (green). Side chains of the two oncogenic mutation sites found at this interface, M918 from the EF loop and E768 on the αC helix, are shown. The side chains of the three chosen 'dimer-deficient' mutation sites—Q910 at the end of the activation loop, P766R from the αC helix and P957R from the αG helix—are shown. (c) SDS-PAGE of samples from affinity purification of the 'dimer-deficient' RET-KD double mutants. B = resin bound GST-RET-KD; E = eluted RET-KD. (d) *In vitro* ADP-coupled kinase activity assay of RET-KD mutants with polyE$_4$Y peptide substrate. The experiment was carried out in triplicate and the graph shows average values ± standard error of the mean (SEM). Data table shows catalytic values based on fitting to the data with Prism (GraphPad).

### 5.5.2 Dimer-deficient pRET-KD$_{P766R/Q910R}$ crystallisation and data collection

I then began crystallisation screening of the dimer-deficient mutants. Initially I was using phosphorylated protein with AMPPNP, since the phosphorylated dimer-deficient RET-KDs could be concentrated up to 5 mg/ml without a significant loss of material,

whereas the solubility of the non-phosphorylated protein was poor above 2 mg/ml. Using phosphorylated material would most likely produce an active conformation of the RET-KD in the crystal, further promoted by the binding of the ATP-analogue AMPPNP. This strategy would therefore not achieve the eventual goal of observing an inactive conformation of the RET-KD, but would determine whether the mutations could induce a different arrangement of the RET-KD in the crystal, with an accessible active site. A crystal structure of the RET-KD without the previously observed *trans*-inhibited dimer, even in an active conformation would reveal whether the previously observed arrangement of the structural elements involved in the dimer interface are at all due to crystal packing effects. It may also allow the RET-KD to adopt a different conformation. Should crystallisation of the dimer-deficient RET-KDs lead to a different lattice arrangement being observed the mutants could then be used in screening known inactive conformation binding inhibitors such as sorafenib, once improvements in the solubility of the unphosphorylated protein had been achieved.

To test whether these mutants could crystallise in the standard sodium formate condition, used for crystallisation of wild type RET-KD, I set up a small screen around the 2 M sodium formate and 0.1 M sodium citrate pH 5.5 condition, alongside screening using a range of commercial kits. Following seeding hexagonal crystals of the P766R/Q910R mutant but not the P766R/P957R mutant were produced in this condition.

Crystals of pRET-KD$_{P766R/Q910R}$ with AMPPNP were also obtained in the first round of crystallisation screening in two non-formate based conditions. Both contained ammonium sulphate and used sodium citrate as a buffer (pH 4 and pH 5.6). These conditions were very similar to the condition that produced the low resolution crystals of pRET-KD:sunitinib (section 5.3.1), and these crystals were also hexagonal (Figure 5.14a). Following production of larger crystals in a larger drop size, the crystals were harvested, cryoprotected in mother liquor plus 25% glycerol, and flash frozen. Crystal X-ray diffraction screening was carried out at Diamond Light Source. The best diffraction obtained went to a resolution of around 7.5 Å. The data could not be indexed from the test shots and therefore data was collected over 180° with a 1° oscillation (Figure 5.14). The dataset images showed evidence of multiple lattices in some directions, although one lattice dominated the diffraction pattern.

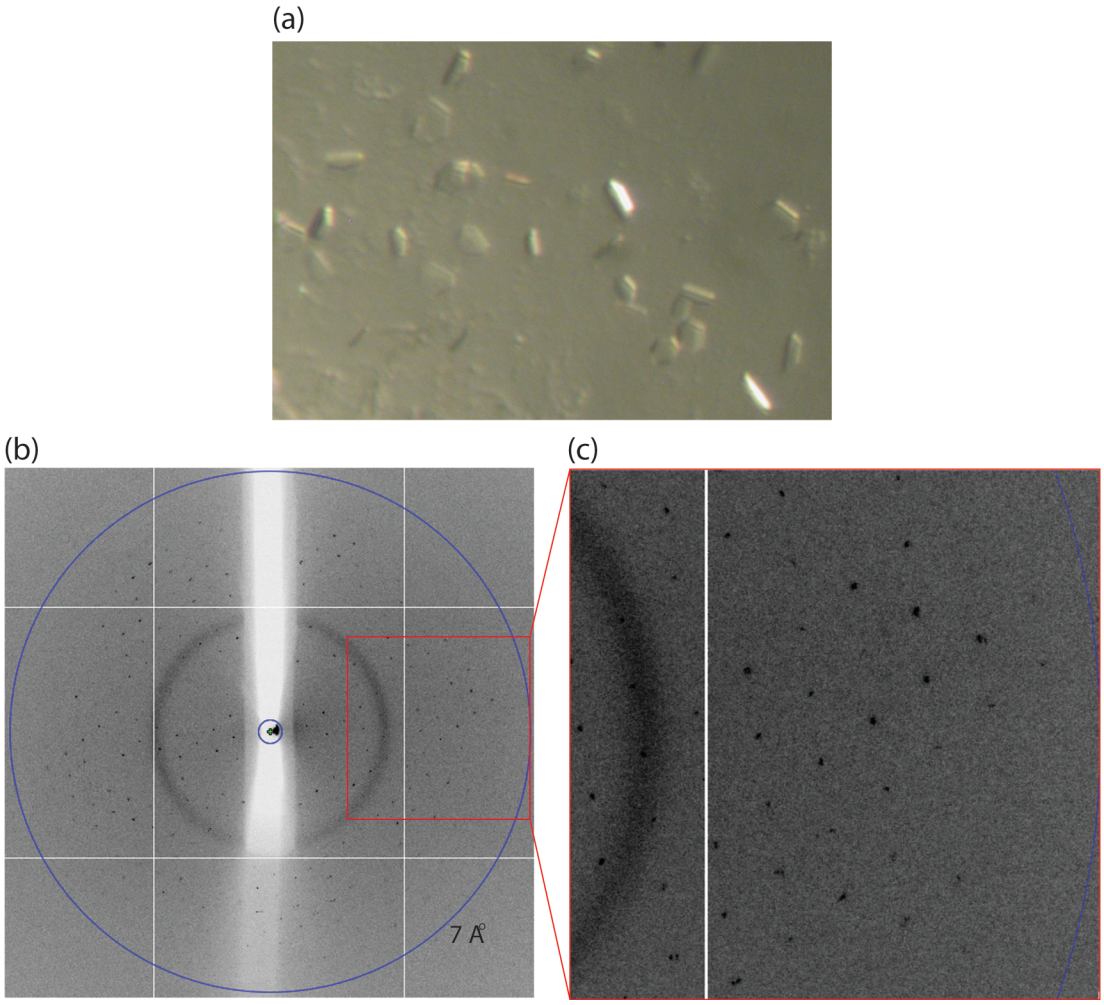



**Figure 5.14 pRET-KD$_{P766R/Q910R}$ with AMPPNP crystals and diffraction image**

(a) Hexagonal crystals of pRET-KD$_{P766R/Q910R}$ with AMPPNP. (b) & (c) Diffraction image from the best pRET-KD$_{P766R/Q910R}$ dataset collected at Diamond Light Source I-03.

### 5.5.3 pRET-KD$_{P766R/Q910R}$ diffraction data processing

Despite the low resolution the collected pRET-KD$_{P766R/Q910R}$ diffraction dataset was processed in the hope of determining the arrangement of the RET-KD molecules in the crystal lattice, in order to determine whether the double mutation had successfully disrupted the crystallographic dimer. The dataset was indexed, integrated and scaled and averaged using XDS (Kabsch, 2010) since MOSFLM failed to index the data. Data statistics, including the refined cell dimensions are detailed in Table 5.5. POINTLESS was used to confirm the correct spacegroup, which was P6$_3$22. Cell content analysis in

CCP4 indicated that there was 1 RET molecule in the asymmetric unit with a 57% solvent content.

Given the low resolution and poor data quality the structure could not be solved by standard molecular replacement using MOLREP (Vagin and Teplyakov, 1997) or PHASER (McCoy *et al.*, 2007) using two different pRET-KD structures. Even separating the search model into the N-lobe (residues 717–803) and C-lobe (848–1005), did not lead to a solution with a reasonable Z score.

However a 3.3 Å RET-KD structure had already been determined in the lab with the same space group and similar unit cell (a = 97.736, b = 97.736, c = 144.982, $\alpha = \beta = 90.00°$, $\gamma = 120.00°$). Since all we wanted to know at this stage was whether the double mutation had successfully disrupted dimer formation, we simply combined the protein coordinates from this structure with the scaled pRET-KD$_{P766R/Q910R}$ data, having altered the cell parameters to match those of pRET-KD$_{P766R/Q910R}$. This was put through ten rounds of restrained refinement with REFMAC. The results are detailed in Table 5.6, with a final $R_{free}$ of 43%. This was sufficient to confirm the overall arrangement of the RET-KD molecules in the crystal, although the data was certainly not good enough for any further interpretations to be made.

| **Data collection** | |
|---|---|
| **X-ray source** | Diamond Light Source I-03 |
| **Wavelength / Å** | 0.9800 |
| **Space group** | $P6_322$ |
| **Unit cell dimensions** | |
| **a, b, c / Å** | 96.37, 96.37, 141.91 |
| **α, β, γ / °** | 90, 120, 90 |
| **Resolution / Å** | 142.57–7.48 (7.88–7.48) |
| **Total reflections** | 7019 (1038) |
| **Unique reflections** | 649 (88) |
| **Redundancy** | 10.8 (11.8) |
| **Completeness / %** | 100.0 (100.0) |
| **I/σ** | 9.2 (4.2) |
| **$R_{merge}$ / %** | 19.3 (59.4) |
| **$R_{p.i.m.}$ / %** | 6.3 (18.0) |

**Table 5.5 Data collection statistics for pRET-KD$_{P766R/Q910R}$ with AMPPNP**

Values for the outermost shell are shown in parenthesises.

| Refinement statistics | |
|---|---|
| **Resolution / Å** | 19.70–7.48 |
| **Completeness / %** | 100 |
| **No. of reflections** | 694 |
| **$R_{work}$ / %** | 30.68 |
| **$R_{free}$ / %** | 42.98 |
| **Correlation coefficient (Fo-Fc) / %** | 79.5 |
| **Total protein atoms** | 2195 |
| **Total ligand atoms** | 21 |
| **Total solvent atoms** | 3 |
| **Mean B-factor / Å²** | 170.7 |
| **Ramachandran Plot / %** | |
| **Favoured** | 90.64 |
| **Allowed** | 7.12 |
| **Outliers** | 2.25 |
| **RMSD from ideal values** | |
| **Bonds / Å** | 0.008 |
| **Angles / °** | 1.138 |

**Table 5.6 Refinement statistics for pRET-KD$_{P766R/Q910R}$ with AMPPNP**

The RET-KDs in this lattice still form the head-to-tail dimer (Figure 5.15). As for the P766R mutation alone, the P766R/Q910R double mutant dimer shows a 20° rotation of the monomers relative to each other compared to the dimer in the wild type RET-KD structure. The P766R/Q910R mutation is therefore not sufficient to prevent head-to-tail dimer formation during crystallisation.

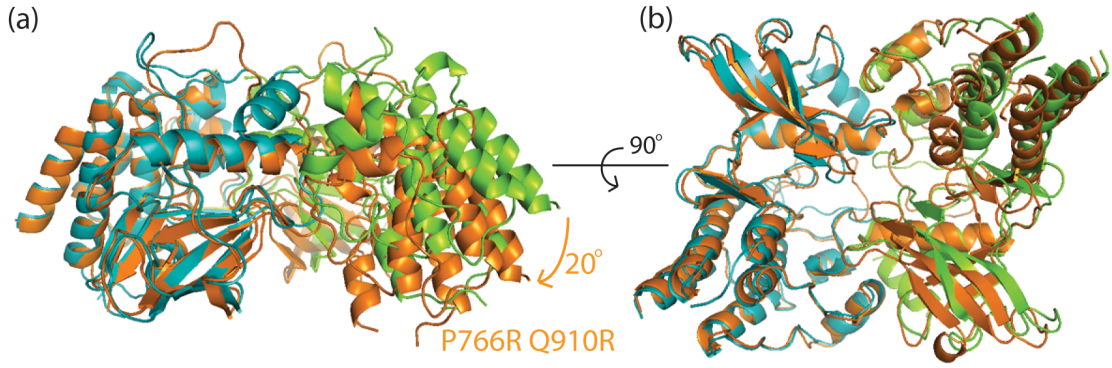



**Figure 5.15 pRET-KD$_{P766R/Q910R}$ crystals still contain the symmetric dimer**

(a) Dimer of pRET-KD:sunitinib structure (teal and green ribbon), with superposition of one monomer of the pRET-KD$_{P766R/Q910R}$ crystallographic dimer (orange ribbon), showing retention of the head-to-tail dimer. A small rotation of the relative position of the two monomers has occurred in the mutant, opening up the activation loop region slightly. This change is likely a reflection of the different space group.

### 5.5.4 Dimer-deficient pRET-KD$_{P766R/R957R}$ further crystallisation screening

Initial screening of pRET-KD$_{P766R/P957R}$ with AMPPNP did not yield any crystals. This included the sodium formate/sodium citrate and ammonium sulphate/sodium citrate conditions, which was encouraging since it means this double mutant is unable to crystallise in the same way as the wild type or pRET-KD$_{P766R/Q910R}$ with AMPPNP. Therefore the P766R/P957R mutation is likely truly 'dimer-deficient'.

To try and encourage crystallisation the pRET-KD$_{P766R/P957R}$ was subjected to a further purification step. Size exclusion chromatography was carried out with a high salt concentration in the buffer (400 mM). The protein eluted almost exclusively as a monomer (Figure 5.16a), unlike the wild type protein, which showed a much higher level of aggregation (though at 300 rather than 400 mM NaCl, Figure 5.2). Following addition of AMPPNP and concentration to 5 mg/ml, DLS was carried out, which confirmed the protein was monomeric at the concentration used in crystal trials (Figure 5.16c).

Using the size exclusion purified material I tried co-crystallisation with PP1 and sunitinib as well as AMPPNP, so see whether a smaller molecule in the nucleotide pocket would encourage crystallisation. This was unsuccessful; so one further modification was made. I tried co-crystallisation with AMPPNP and a 16 residue peptide of the RET activation loop (AL-peptide; residues S896–G911: SRDVYEEDSYVKRSQG). Screening of this did produce apparently crystalline objects in a PEG monomethyl ether 5000 based condition. These disintegrated during harvesting and cryoprotection however, and an initial round of optimisation, varying the PEG concentration and the pH with an increased drop size, did not result in improved crystal stability. Since better crystals were produced using the wild type protein with this peptide (see section 5.6), further attempts at crystal optimisation were not carried out.

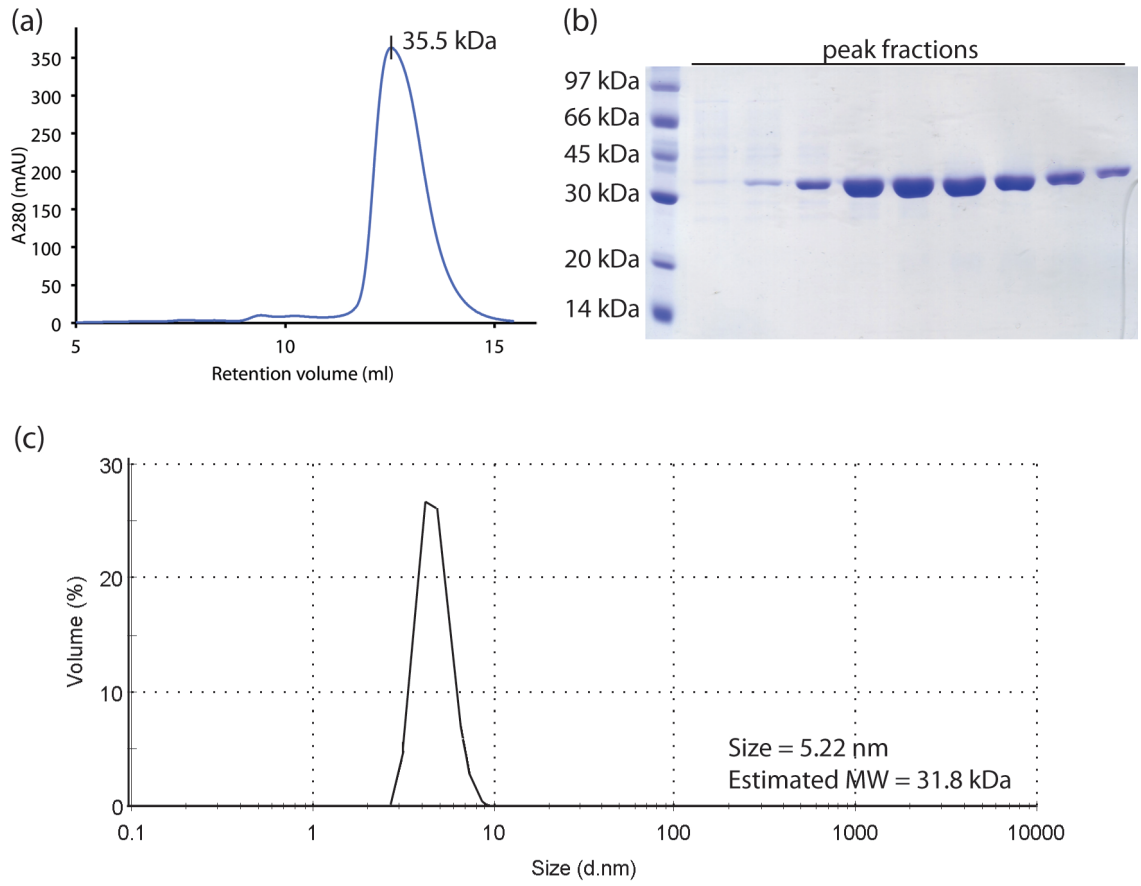



**Figure 5.16 pRET-KD$_{P766R/P957R}$ further purification by size exclusion chromatography**

(a) Size exclusion of pRET-KD$_{P766R/P957R}$ on a Superdex 75 column with 400 mM NaCl in the buffer. (b) Coomassie stained reducing SDS PAGE of peak fractions from size exclusion shows purity of protein. (c) Dynamic light scattering results of size exclusion purified pRET-KD$_{P766R/P957R}$ with AMPPNP showing the protein in monodisperse in solution.

## 5.6 Towards a crystal structure of the RET-KD with a substrate peptide

Another way of obtaining a different crystal form of the RET-KD, this time specifically to look further at the active conformation of the kinase, would be to crystallise RET in complex with a substrate or substrate peptide and ATP-analogue. Since the previously observed the crystallographic dimer is incompatible with substrate binding (see section 1.7.2) this would require an alternative arrangement of the RET-KD molecules in the crystal. Were this successful the resulting crystal structure would reveal the details of substrate interaction with the RET-KD as well as the ATP compatible arrangement of

the gly-rich loop. This might provide insights into the oncogenic mutations near the substrate binding site, such as M918T, as well as providing a basis for development of substrate-mimetic inhibitors.

I screened pRET-KD with a three fold molar excess of both AMPPNP and the activation loop peptide (AL-peptide) and initially obtained crystals in two conditions (top two in Table 5.7). The AL-peptide has a very low pH, so in order to have greater control over the pH in the crystallisation trials the peptide was brought to pH 7 with sodium hydroxide. Further screening with fresh protein was then carried out in PEG based kits. The conditions that produced crystals, all with plate morphology, are listed in Table 5.7. Common themes in the conditions are the absence of salt and the presence of a low to medium molecular weight PEG, often with a citrate-containing buffer (pH 5.5–8). It is assumed that the AL-peptide is bound to the pRET-KD in the crystal since pRET-KD does not crystallise in these conditions without peptide present, however until the crystal structure is solved this cannot be definitively proven.

| Kit reference | Buffer | Precipitant | Additive |
|---|---|---|---|
| **Index H10** | - | 20% (w/v) PEG 3350 | 0.2 M sodium citrate tribasic |
| **PEG Rx F10** | 0.2 M ammonium citrate pH 7 & 0.1 M imidazole pH 7 | 20% (w/v) PEG 2000mme | - |
| **PEG/Ion E7** | 0.1 M di-sodium malonate pH 6 | 12% PEG 3350 | - |
| **PEG/Ion F2** | 8% tacismate pH 6 | 20% PEG 3350 | - |
| **PEG/Ion F3** | 12% tacismate pH 7 | 12% PEG 3350 | - |
| **PEG/Ion F5** | 4% tacismate pH 8 | 12% PEG 3350 | - |
| **PEG Rx B5** | 0.1 M sodium citrate pH 5.5 | 22% PEG 1000 | - |
| **PEG Rx F2** | 0.1 M sodium citrate pH 5.5 | 20% PEG 1000 | 0.1 M $LiSO_4$ |
| **PEG Rx G2** | 0.1 M imidazole pH 7 & 2% tacismate pH 7 | 8% PEG 3350 | 5% 2-propanol |

**Table 5.7 pRET-KD + AMPPNP + AL-peptide crystallisation conditions**

Index, PEG Rx and PEG/Ion are crystallisation screening kits from Hampton Research. For crystallisation screening 3.7 mg/ml pRET-KD + AMPPNP + ALpeptide was mixed with well solution at a 1:1 ratio in sitting drops and incubated at 20 °C.

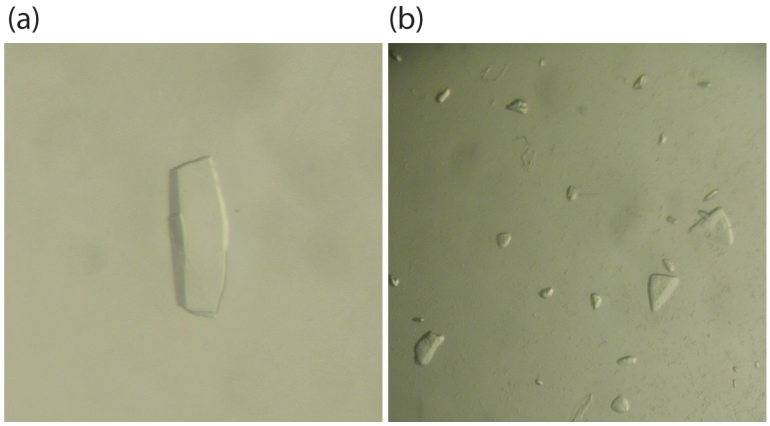



**Figure 5.17 Crystals of pRET-KD with AMPPNP and AL-peptide**

(a) and (b) Images of pRET-KD + AMPPNP + AL-peptide plate crystals grown in 18–20% (w/v) PEG 3350, 0.2 M sodium citrate tribasic.

To produce larger crystals than obtained in initial crystallisation screening, screens with larger drops based around some of the hit conditions were set up. Plates of up to 120 μm were produced on the first round of optimisation (Figure 5.17). Several of these were cryoprotected with well solution plus 25% glycerol, harvested, and flash frozen in liquid nitrogen. Testing crystals for X-ray diffraction was carried out at Diamond Light Source on a microfocus beamline (I-24). Many of the crystals contained multiple lattices, and a smeared spot shape was common. The resolution was mostly between 6–8 Å, although one crystal diffracted to 4.3 Å, but only in one direction. The crystal that produced the best spot shape only diffracted to 8 Å. A dataset from this crystal was collected in order to determine the space group and unit cell dimensions.

Since it was collected using a Pilatus detector the 8 Å dataset was processed using D*TREK (Pflugrath, 1999). The space group was found to be C2 with unit cell dimensions as detailed in Table 5.8. The Matthews coefficient indicated that there were between 3–5 RET-KD molecules in the asymmetric unit. This is a different crystal form to those obtained for RET-KD previously (Table 5.8). Therefore if these new crystallisation conditions can be optimised to yield single crystals with better diffraction, the resulting structure could yield new information about RET-KD's regulation and interaction with substrate. Unfortunately since these crystals were obtained late in the course of this thesis, I was not able to perform this further crystal optimisation myself.

| Crystal | | pRET-KD: adenosine | Non-phospho. RET-KD:cAMP | pRET-KD$_{P766R/Q910R}$: AMPPNP | pRET-KD: AMPPNP: AL-peptide |
|---|---|---|---|---|---|
| **Crystallisation Condition** | | 2 M sodium formate, 0.1 M sodium citrate pH 5.5 | 2 M sodium formate, 0.1 M sodium citrate pH 5.5 | 1 M ammonium sulphate, 0.1 M sodium citrate pH 5 | 14% PEG 2000mme, 0.2 M sodium citrate tribasic |
| **Highest observed resolution** | | 1.8 Å | 2 Å | 7.5 Å | anisotropic 4.3/8 Å |
| **Space group** | | C2 | $P2_1$ | $P6_322$ | C2 |
| **Cell lengths (Å)** | a | 73.1 | 50.4 | 96.4 | 97.2 |
| | b | 69.0 | 80.2 | 96.4 | 171.5 |
| | c | 78.6 | 79.7 | 141.9 | 85.4 |
| **Cell angles** | α | 90° | 90° | 90° | 90° |
| | β | 101.8° | 100.9° | 90° | 113.7° |
| | γ | 90° | 90° | 120° | 90° |
| **Molecules in the asymmetric unit** | | 1 | 2 | 1 | 4 (3–5) |
| **Cell volume (Å$^2$)** | | 388469 | 317171 | 1141372 | 1304932 |

**Table 5.8 RET KD crystal forms**

This table details examples of the 4 crystal forms of the RET-KD that have so far been obtained. Three of these are examples from this thesis and the non-phosphorylated (non-phospho.) RET-KD was crystallised by Phillip Knowles (Knowles *et al.*, 2006). All published RET-KD crystal structures are from crystals produced in a very similar condition to pRET-KD:adenosine, have C2 symmetry and exhibit very similar unit cell dimensions, except the non-phosphorylated RET-KD, which shows the same RET-KD dimer despite the slightly different lattice.

## 5.7 Conclusions

In this chapter I have presented a 2.2 Å crystal structure of the RET-KD bound to the clinically relevant RET-inhibitor sunitinib. This showed sunitinib binding the active *DFG-in* conformation of the RET-KD, which is unlike sunitinib's interaction with KIT, where it preferentially binds the inactive *DFG-out* conformation of the KIT-KD.

I have also presented a high resolution (1.8 Å) crystal structure of the RET-KD bound to adenosine, which revealed some previously unseen features of the RET-KD structure, including an alternative conformation of the flexible gly-rich loop. A close look at this crystal structure also generated a new hypothesis for the mechanism of action of a RET-

KD oncogenic mutation, which is being tested in mammalian cells in the Structural Biology Laboratory.

Two mutations (P766R/P957R), which together disrupt the formation of the crystallographic *trans*-inhibited dimer, have been presented. This 'dimer-deficient' double mutant could be used as a tool for testing the *trans*-inhibited dimer hypothesis in cells. Unfortunately crystallisation screening of this mutant was unsuccessful and therefore I have not been able to assess the role of the crystallographic dimer in the structure of the RET-KD in the crystal.

Finally I have presented a method for the crystallisation of RET-KD with a substrate peptide. These crystals are a new RET-KD crystal form, and if further optimisation improved the resolution of the crystals, the resulting structure could reveal new insights into RET-KD function and substrate interaction.

# Chapter 6. Discussion

Despite well established physiological and pathophysiological roles for RET in GFL/GFRα signalling, a molecular description of RET activation and oncogenic deregulation is poorly understood. Similarly, an integrated understanding of the relationship between the extracellular and intracellular parts of RET during ligand-dependent activation is completely lacking. In this thesis I have sought to determine the architecture of the ligand-binding RET-ECD alone and in complex with ligand/co-receptor with a view to understanding how RET binds ligand/co-receptor and how this engagement might lead to RET intracellular tyrosine kinase activation. I have also sought to develop our understanding of the regulation and chemical inhibition of the RET-KD by producing new crystal structures of the RET-KD. In this chapter I will discuss the implications of the RET-ECD and GDNF/GFRα1/RET-ECD ternary complex architecture presented in this thesis for understanding published functional data and for understanding disease-associated RET mutations. I will also briefly discuss the limitations of our current understanding of RET regulation and outline some possible strategies to develop this going forward.

## 6.1 A SAXS-derived model of the RET ECD

The SAXS-derived model of the subdomain arrangement defining the overall zRET-ECD architecture shows: a good fit for the clamshell CLD1–CLD2 arrangement, as observed in the hRET-CLD(1–2) crystal structure; a near-linear (~150°) calcium-rigidified CLD2–CLD3 arrangement; a 100° bent CLD3–CLD4 arrangement, which is to my knowledge a novel tandem cadherin domain arrangement; and the CRD, at least in part, folded back on CLD4. This SAXS-derived model provides a first look at the overall RET-ECD architecture and, when combined with other structural and functional data, provides some insights into a possible mechanism of RET regulation and how RET-ECD Hirschsprung's mutations may impact on RET folding.

### 6.1.1 The role of calcium in defining the CLD2–CLD3 junction

The collapse of the near-linear CLD2–CLD3 arrangement in the absence of calcium, as observed by SAXS, indicates that like EC–EC junctions in classical cadherins, calcium is required for the end-to-end CLD2–CLD3 arrangement (Haussinger *et al.*, 2002; Pokutta *et al.*, 1994). Highlighting this in cadherins, a recent crystal structure of a calcium-free tandem cadherin domain array from T-cadherin showed its EC(1–2) domains adopting a clamshell arrangement with EC1 folded back onto EC2, similar to the constitutively calcium-free, disulphide-tethered, RET CLD1–CLD2 arrangement (Ciatto *et al.*, 2010; Kjaer *et al.*, 2010). In the T-cadherin case calcium was chelated from the EC1–EC2 junction by the citrate present in the crystallisation condition, leading to the collapse of the previously characterised near-linear calcium-rigidified EC1-EC2 arrangement (Ciatto *et al.*, 2010). In this study my calcium-free zRET-ECD samples were purified in the absence of calcium but no attempt was made to chelate calcium from the protein. The inability of zRET-ECD prepared in the absence of calcium to bind zGFRα1a/zGDNF, unless calcium was added to the buffer, combined with the SAXS data showing the collapse of the near-linear CLD2–CLD3 arrangement, strongly suggests that there was no calcium associated with the zRET-ECD purified in the absence of calcium. Since calcium is required for RET-ECD production (van Weering *et al.*, 1998), this indicates that calcium dissociated from the CLD2–CLD3 junction during purification in the absence of calcium in the purification buffers. zRET-ECD would therefore require calcium to be present in the extracellular space for maintenance of its architecture and functionality. This is consistent with previous data showing calcium is required for hRET signalling (Anders *et al.*, 2001; Nozaki *et al.*, 1998). The affinity of the three calcium-binding sites at the RET CLD2–CLD3 junction has not been measured. In classical cadherins the affinity of the calcium-binding sites varies from the micromolar to millimolar range and N-cadherin has been shown to be sensitive to the extracellular calcium concentration on a rapid timescale and therefore has been proposed to be regulated by calcium (Kim *et al.*, 2011). RET could therefore be similarly reversibly regulated by extracellular calcium levels although this has not been shown.

### 6.1.2 The CLD3–CLD4 junction, an unusual tandem cadherin domain arrangement

The SAXS-derived zRET-ECD model shows a 100° bend angle between CLD3 and CLD4, which was apparently maintained in the absence of calcium. This is distinct from the CLD1–CLD2 clamshell arrangement and also from the 80° bend angle observed in the first crystal structure of a constitutively calcium-free tandem cadherin domain array (EC(2–3)) from non-classical *Drosophila* N-cadherin (DN-cadherin) (Jin *et al.*, 2012). It is also distinct from the calcium-bound EC–EC junctions observed in crystal structures of classical cadherins which show interdomain angles of ~140–160° in C-cadherin, ~145–155° in N-cadherin and ~130–170° in E-cadherin (Boggon *et al.*, 2002b; Harrison *et al.*, 2011).

The structural basis for the observed CLD3–CLD4 arrangement will not be clear until the CLD(1–4), or another CLD3–CLD4 containing construct, crystal structure is solved. Progress towards such a CLD(1–4) crystal structure was described in Chapter 2. Sequence alignment of RET CLDs to classical cadherin ECs identified a long hydrophobic 20–30 amino acid linker between predicted CLD3 βG and CLD4 βA, which is relatively well conserved amongst vertebrate RETs (Anders *et al.*, 2001) (alignment is shown in Figure 1.7). Given that it is predominantly hydrophobic, this linker is presumably mostly buried at the CLD3–CLD4 junction. There is another plausible interpretation of the arrangement of structural elements at the CLD3–CLD4 junction however. The CLD3–CLD4 'linker' contains a conserved LXF motif (Figure 6.1), which defines the start of CLD1 and CLD2, with the CLD3 start defined by the cadherin-conserved PXF motif. The phenylalanine from this motif packs into the core of the β-barrel fold, stabilising βA at the start of each CLD or EC. Secondary structure prediction in PHYRE2 predicts with high confidence (9/9) that a large proportion of the CLD3–CLD4 'linker' forms a β-strand. Due to these two features an alternative hypothesis is that this 'linker' β-strand is actually CLD4 βA, with the previously predicted βA actually βB. The previously predicted βB might then form a long, partly basic and potentially helical linker between βB and βC at the CLD3–CLD4 junction, in line with secondary structure prediction for this region. This linker would contain the conserved potentially disulphide-constrained CV/IENC loop (Figure 6.1). All the

previously proposed CLD4 β-strands apart from βB are predicted to be β-strands by secondary structure prediction, and align well with cadherin consensus motifs. Resolution of this awaits a crystal structure, but the CLD3–CLD4 junction is evidently a divergent cadherin junction by sequence alignment, resulting in a novel cadherin domain arrangement as observed by SAXS.

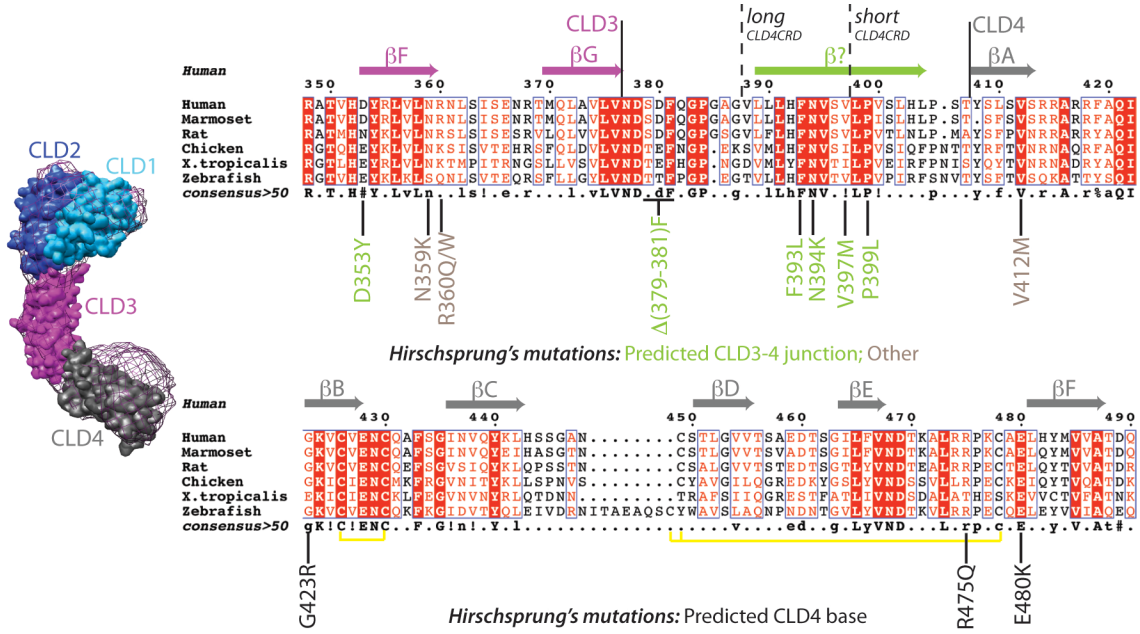



**Figure 6.1 Hirschsprung's mutations at the CLD3–CLD4 and CLD4–CRD junctions**

Vertebrate RET CLD3–CLD4 junction sequence alignment showing the predicted β-strands based on cadherin alignment (CLD3, magenta; CLD4, grey) and corroborated by secondary structure prediction in PHYRE2, except CLD4 βB, which is predicted to be helical. The secondary structure predicted β-strand in the proposed CLD3–CLD4 junction is coloured green. Yellow lines show the predicted disulphide pattern. Hirschsprung's-associated mutations are annotated beneath the alignment. Most of the mutations in this region of RET are located at the predicted CLD3–CLD4 junction or the base of CLD4, potentially engaging the CRD. The N-termini of the two zRET-CLD4CRD constructs (long and short) produced in this thesis are annotated above the alignment.

The importance of the CLD3–CLD4 junction in the folding (or function) of the RET-ECD is highlighted by the presence of a cluster of Hirschsprung's mutations at the predicted junction (Figure 6.1). One of these is an N-linked glycosylation site mutation (N394K). D353Y and Δ(379–381)F result in the loss of acidic residues from the junction. F393L (from the putative LXF motif), V397M and P399L are at the heart of the previously predicted CLD3–CLD4 junction and F393L has been shown to

kinetically compromise hRET-ECD folding, but not affect GDNF/GFRα1 binding (Kjaer *et al.*, 2010; Kjaer and Ibanez, 2003b).

### 6.1.3 Structural characterisation of the CRD

In both the SAXS-based model of the zRET-ECD and the presented interpretation of the mTC 3D reconstruction the CRD is, at least partly, folded back and packed against CLD4. Unlike the CLDs the CRD fold cannot be readily predicted based on sequence or predicted secondary structure homology and unfortunately I have not been successful at obtaining crystals of a zRET-ECD construct containing the CRD. In the absence of X-ray crystallography data, circular dichroism experiments indicated the CRD largely lacks regular secondary structural elements although it may have some proportion of β-strands. The confidence of β-strand prediction by circular dichroism is poorer than α-helix prediction however as the circular dichroism signal for β-strands is weak (Martin and Schilstra, 2008). The CRD fold is likely dominated by the 7–8 disulphides and a number of the CRD cysteine residues are sites of Hirschsprung's mutations, with mutation resulting in impaired folding. The other CRD-associated Hirschsprung's mutations both at the predicted base of CLD4 (Figure 6.1) and within the CRD (Figure 6.2) commonly result in loss or gain of a charged residue in this largely polar domain.

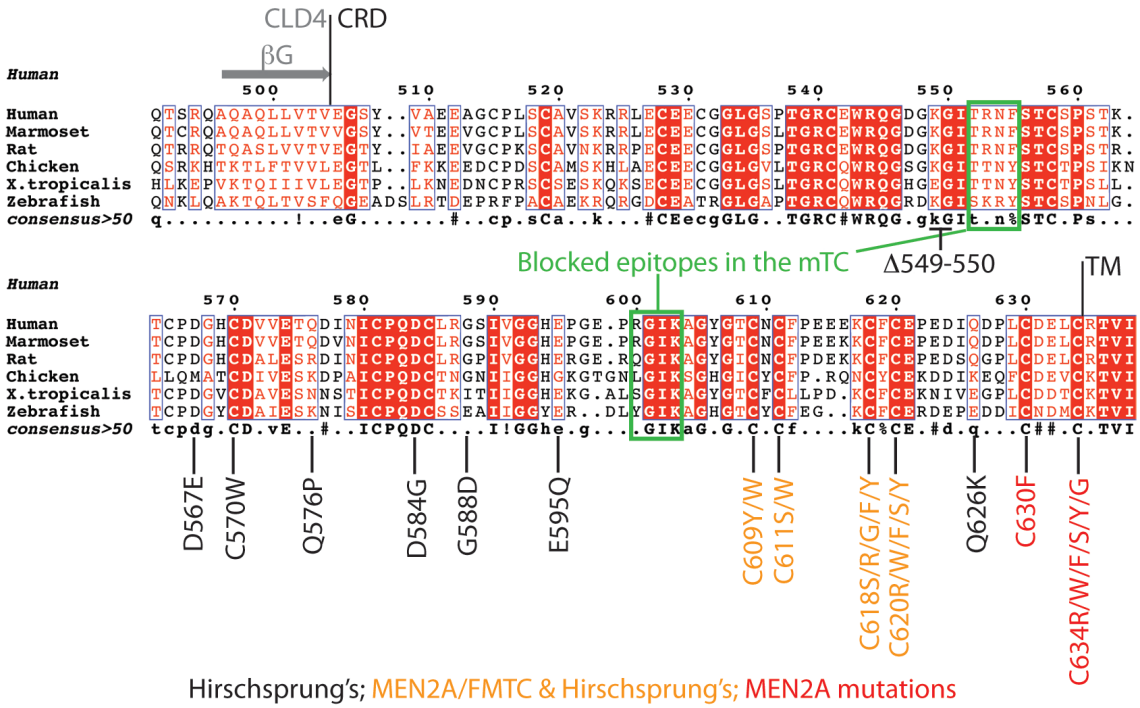



**Figure 6.2 The CRD associated disease mutations and role in ternary complex formation**

Vertebrate RET-CRD sequence alignment with the known Hirschsprung's and MEN2-associated mutations indicated beneath the alignment. The epitopes recognised by two antibodies that bind RET alone but not in the context of the mTC are marked by green boxes.

### 6.1.4 A potential interaction between CLD3 and the CRD

When zRET-CLD4CRD is produced alone around 20% of the protein forms disulphide-linked dimers. Such covalent species are not observed when the entire zRET-ECD is purified and implies the presence of an accessible and unpaired cysteine in the zRET-CLD4CRD that is either not accessible and/or always paired in the zRET-ECD. The zRET-CLD4CRD contains an even number of cysteine residues and CLD3 contains no cysteine residues. Therefore the zRET-CLD4CRD construct is unlikely to contain an unpaired cysteine due to removal its usual disulphide partner. Unpublished observations by Dr. Svend Kjær working with human RET-ECD and CLD4CRD proteins revealed that CLD(1–3) influences the arrangement of CLD4CRD. He has found that some antibodies, which recognise epitopes in the CRD, only bind in the context of the whole ECD and not when only CLD4CRD are present. The arrangement of CLD4CRD when they are produced alone is therefore different to their arrangement in the context of the whole RET-ECD. One could speculate that the CRD might not be folded back on CLD4

in the same way, due to the absence of the CLD3-CRD contact. If true this altered CLD4CRD arrangement may be the basis for the observed zRET-CLD4CRD dimers. The changed CLD4CRD arrangement could separate two cysteines normally involved in formation of an intramolecular disulphide, making formation of this bond less likely and resulting in at least a proportion of monomers with unpaired cysteine residues. These can then form intermolecular disulphide bonds resulting in the observed disulphide-linked dimers.

An alternative hypothesis is the lack of CLD3 exposes the top of CLD4, which is predicted to be the location of a short, conserved, potentially disulphide-constrained loop: CV/IENC. It could therefore be these, ordinarily protected, cysteines that are involved in the observed intermolecular disulphides.

The disulphide-linked CLD4CRD dimer is intriguing since some of the oncogenic MEN2A mutations in RET occur at cysteine residues in the CLD, resulting in constitutive disulphide-linked dimers in the plasma membrane. It is possible I have therefore inadvertently trapped a species like one of the MEN2A mutants. Comparative mass spectrometry of trypsin-digested monomer versus dimer was attempted to see whether the intermolecular disulphide could be identified by the presence of a unique disulphide-linked peptide in the dimer. However the C-terminal peptide of the CRD derived from trypsin-digestion, where the interesting MEN2A cysteines reside, was too large for analysis by mass spectrometry with the available setup. Use of a different protease to digest the CLD4CRD monomer and dimer to produce smaller peptides of this region could help if this were to be repeated.

## 6.2 The architecture of a mammalian RET/ligand/co-receptor ternary complex

In Chapter 3 I presented a 24 Å resolution negative stain EM 3D reconstruction of a mammalian RET ternary complex. An interpretation of the arrangement of the RET-ECD, GFRα1 and GDNF in the RET ternary complex was generated based on fitting of the mTC components with known or modelled structure into the EM map. Several assumptions were made before fitting as discussed in Chapter 3. The hand of the EM

map also had to be assigned and was chosen based on consistency of the resulting model with previous RET-GFRα1 interaction site data. The resulting selected interpretation has the 2-fold symmetric GDNF/GFRα1 complex at the heart of the mTC, with the two RET-ECD molecules wrapped around the outside of the GDNF/GFRα1 complex making contacts with both GDNF and GFRα1, resulting in an overall two-fold symmetric arrangement. Here I will discuss the implications of this current interpretation.

### 6.2.1 RET activation by ligand/co-receptor engagement

The current interpretation of the mTC 3D reconstruction reveals a central role for the ligand/co-receptor complex in organising and mediating RET dimerisation. Multiple interactions between the RET-ECD and ligand/co-receptor result in a ligand/co-receptor-defined stable arrangement of the two RET-ECDs in the complex, presumably positioned to optimally arrange the membrane proximal regions of the RET-ECD for activation of the intracellular kinase domain in the full-length receptor.

In the current interpretation the membrane-proximal CRDs are drawn towards each other at the base of the mTC, interacting with opposing faces of GDNF. Part of the CRD is potentially also extending down into the base region of the complex towards the 'membrane'. Although since the crystal structure of the CRD is unknown and therefore cannot be fitted into the EM map this is purely speculative. The central location of the CRD in the mTC, packed between CLD4 and GDNF, is consistent with unpublished data from Dr. Svend Kjær. He has observed that two antibodies, which recognise different epitopes in the CRD, do not bind hRET-ECD when present within the mTC. This indicates that these epitopes are buried in the complex (the epitopes are shown in Figure 6.2). Juxtaposed CRDs extending towards the membrane in the base region of the 3D reconstruction is consistent with the oncogenic activating potential of MEN2A cysteine mutations towards the C-terminus of the CRD. The oncogenic disulphide-linked RET dimers that result from these cysteine mutations presumably recapitulate and stabilise the arrangement of the membrane-proximal region of the RET-ECD in the wild type 'active' RET/ligand/co-receptor complex independent of ligand/co-receptor binding, through covalent CRD:CRD interactions.

The RET-CLD(1–4) arrangement from the zRET-ECD SAXS analysis has been fitted as a rigid body in the EM map and there is no evidence of large conformational changes in this portion of the zRET-ECD when bound to ligand/co-receptor. Indeed given that CLD(1–2) arrangement is disulphide constrained and the CLD2–CLD3 junction is calcium-rigidified, large conformational changes in CLD(1–3) are not expected. The rigidity of the CLD(3–4) arrangement is unknown, but it appears to be maintained in this current interpretation of the mTC 3D reconstruction. At the current resolution of the EM map localised conformational changes may exist within CLD(1–4) but cannot be determined with any certainty.

Allosteric changes such as the movement of part of the CRD away from CLD4 upon ligand/co-receptor engagement may conceivably occur. For example, if the CRD does extend into the base region of mTC this would contrast with its location against CLD4 as observed by SAXS of the ECD alone. However greater structural understanding of this domain and better resolution EM maps are required to prove this hypothesis. We note that precedents have been described previously for the EGFR, where binding of EGF ligand leads to a conformational change resulting in the exposure of a dimerising arm from the receptor that is otherwise buried in the absence of ligand (Ogiso *et al.*, 2002).

### 6.2.2 Differential signalling by RET in response to different ligands

The arrangement of the RET-ECD wrapped around the ligand/co-receptor complex is consistent with the hypothesis of Goldman and colleagues that RET is able to detect and respond to the different bend angles observed in GFL dimers (Parkash and Goldman, 2009; Parkash *et al.*, 2008). The different bend angles observed in GDNF and Artemin result in the Artemin/GFRα3 complex being more linear and longer than the GDNF/GFRα1 complex (shown in Figure 1.5). If RET-ECD binds Artemin/GFRα3 in the same manner as GDNF/GFRα1 in the current interpretation of the mTC this would result in RET-ECDs being flatter with respect to the membrane. This means that presumably either the bend angle between the RET transmembrane domain (TM) and CLD4 is different—135° for RET bound to GDNF/GFRα1 versus ~115° for RET bound to Artemin/GFRα3, based on the TM being perpendicular in the membrane—

requiring the CRD to be flexible. Or alternatively the CLD4-TM angle is maintained by a rigid CRD and the arrangement of the two TMs is changed with respect to each other. The latter could result in a changed arrangement of the RET-ICDs, which may be the basis of the differential downstream signalling initiated by RET in response to the different ligands (Parkash and Goldman, 2009).

### 6.2.3 Interactions between RET-ECD and GFRα1/GDNF in the mTC

Previous published studies looking at the interactions between RET and ligand/co-receptor, had the following findings: RET-CLD(1–3) are necessary for GFL/GFRα binding by human RET-ECD and confer species specificity to the interaction. Elements within CLD1 are particularly important, and may influence GFL/GFRα binding indirectly, via mediating a RET:RET interaction that stabilises the GFL/GFRα/RET interaction. RET-CLD4CRD also has a role in GFL/GFRα binding, and may directly interact with GFRα1-D1 and the GDNF N-terminal tail. GFRα-D2D3 contains the major RET binding site although it is unknown where this region binds within RET and GFRα's don't bind RET in the absence of GFLs (these data are summarised in Figure 1.10). It was previously difficult to rationalise all of these data. However the interpretation of the mTC 3D reconstruction presented in this thesis provides a rational basis for several of these findings: CLD(1–3) interacts with GFRα1-D3, explaining the involvement of these domains in the formation of the ternary complex, although this does not explain the species-specificity. The CRD interacts with GDNF and GFRα1-D2 explaining the requirement for both the CRD and GDNF. Examination of these contact regions, predicted by the current interpretation of the 3D reconstruction, revealed several potential *bona fide* RET contact sites on GFRα1 and GDNF, based on the conservation of surface-exposed residues. These proposed binding sites are discussed below, with reference to published data.

#### *6.2.3.1 Proposed RET-CLD(1–3) binding sites on GFRα1-D2D3*

In the current interpretation of the 3D reconstruction the major interaction between GFRα1 and RET is between the very tip of GFRα1-D3 and CLD1 (Figure 6.3). This region of GFRα1-D3, involving the turn between α9 and α10 and the α10 helix,

contains an SGNxxE/DE/D motif, which is conserved amongst all four GFRαs in vertebrates. The pair of glutamate/aspartate residues have previously been shown by mutagenesis studies to be important for the GFRα1/GDNF interaction with RET, with the double alanine mutation reducing RET signalling to 37% of the level with wild type GFRα1 (Parkash *et al.*, 2008). The asparagine has also been previously proposed to be involved in RET binding, since it is surface-exposed and highly conserved; however this has not yet been tested (Wang *et al.*, 2006). The hRET-CLD1 surface facing GFRα contains a number of mammal-conserved basic residues with the potential to interact with these conserved GFRα acidic residues (Figure 6.3). Charge reversal mutations of two of these basic residues (R77D/R144E) reduced the affinity of RET-ECD for immobilised GFRα1/GDNF by ten-fold without impacting on the RET structure (Kjaer *et al.*, 2010).

The 3D reconstruction current interpretation also predicts a possible minor interaction between the loop between α8 and α9 in GFRα1-D3 and the CLD2–CLD3 junction. This GFRα1-D3 loop contains charged residues and a surface exposed hydrophobic residue. However it is not well conserved amongst GFRαs (Figure 6.3) and therefore does not seem a likely to play a major role in the RET:GFRα interaction.

Based on the crystal structure of the hGDNF/rGFRα1-D2D3 complex, Goldman and colleagues identified a number of residues within a charged patch on the surface of the co-receptor, including the previously mentioned acidic residues from the SGNxxEE motif, which when mutated decreased levels of RET autophosphorylation (Parkash *et al.*, 2008). They therefore postulated that these residues were involved in direct RET interaction. These residues are shown in Figure 6.3 and other than the previously discussed pair of acidics they do not appear to make direct contacts with the RET-ECD in this interpretation of the 3D reconstruction. These residues (R190, K194, R197, Q198, K202, R257 and R259) form an SOS (heparin mimic)-binding site in the GFRα1/GDNF crystal structure. It is therefore possible that their mutation in soluble GFRα1-D2D3 leads to decreased RET signalling indirectly by decreasing the affinity of GFRα1-D2D3 for heparin, which may be involved in the capture of GFRα at the membrane. However the authors also state that mutation of these residues reduces RET-binding in their binding studies, although they do not show these data (Parkash *et al.*, 2008). The

involvement of these previously-proposed RET-binding residues within the current interpretation presented in this thesis therefore remains uncertain.

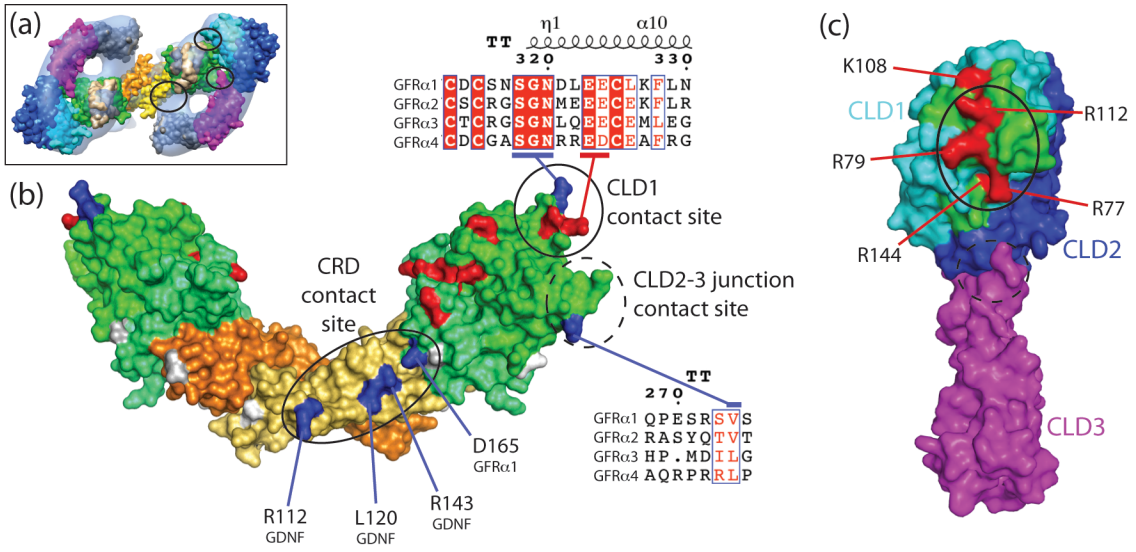



**Figure 6.3 Proposed RET contact sites on GDNF/GFRα1**

(a) The mTC current interpretation of the 3D reconstruction, with the RET GDNF/GFRα1 contact sites circled. The mTC EM map is shown as a transparent blue surface, with fitted RET, GFRα1 and GDNF structures shown in surface representation at 4 Å. (b) Surface view of the GDNF/GFRα1-D2D3 crystal structure showing the proposed RET contact sites from the mTC model. GDNF monomers are coloured yellow and orange, GFRα1-D2D3 green. N-linked glycosylation sites are coloured white and N-linked glycosylation sites in other GFRα's, grey. Residues which have been shown to be important for RET signalling by mutagenesis experiments (Parkash *et al.*, 2008) are coloured red. Surface exposed GFRα/GFL-conserved residues at proposed contact sites are coloured blue. Sequence alignments between human GFRα1–4 for the two proposed GFRα-D3 contact sites are shown, illustrating the conservation of the proposed CLD1 interaction site. (c) Surface view of hRET-CLD(1–3) highlighting the proposed GFRα1-D3 contact sites. Regions at the contact site shown to be important for GFRα1/GDNF binding by homologue scanning mutagenesis (Kjaer and Ibanez, 2003a) are coloured green. Mammalian RET conserved basic residues in these regions are coloured red.

#### *6.2.3.2 Proposed RET-CRD binding site on GDNF*

The current interpretation of the mTC 3D reconstruction predicts an interaction between the CRD and GDNF 'finger 1' and N-terminal tail regions. It also predicts a possible interaction of the CRD with GFRα1-D2 close to the GDNF interface. Surface exposed GFL-conserved residues and a GFRα-conserved residue in this region that could form part of the CRD binding site are shown in Figure 6.3. Residues in these regions have not

been previously been tested for their role in mTC complex stability but experiments can now be undertaken to explore this putative binding interaction.

### 6.2.4 The basis for CLD1 elements conferring species specificity is unclear

The conservation of vertebrate GFRα's and GFLs makes the basis of human/*Xenopus* species specificity in the RET:GFRα1/GDNF interaction, primarily conferred by CLD1 elements, difficult to rationalise in the context of an interaction between CLD1 and the GFRα-D3. Evidently chimeric approaches are limited by the ability of the chimeric proteins to fold properly and such cut-and-paste approaches to protein structure-function studies do not always work.

Class averages from negative stain EM of the zTC showed an apparently similar overall architecture to the mTC, implying the arrangement of the components in the lower vertebrate complex is very similar to the mammalian complex. The absence of the mammal-specific CLD1 elements—the βC'–βC" strand and *cis*-Pro loop—will result in a different CLD1 surface topology in lower vertebrate RETs, although there are still a few basic residues in lower vertebrate RET-CLD1 at the proposed GFRα1-D3 contact site, which could interact with the conserved acidic GFRα1-D3 motif.

## 6.3 Regulation and chemical inhibition of RET-KD

Two new RET-KD crystal structures are presented in Chapter 5 of this thesis. One of these structures has the chemical inhibitor sunitinib bound in the nucleotide-binding pocket, whilst the second structure has adenosine, derived from the hydrolysis of AMP-PNP, bound. The sunitinib-bound structure provides an atomic description of sunitinib interactions with the nucleotide-binding pocket revealing hydrogen-bonding contacts to exclusively mainchain atoms and the disposition of the inhibitor relative to the gatekeeper V804. The high-resolution adenosine bound structure identified two distinct conformers of the gly-rich loop of the RET-KD. One of the conformers is competent to bind ATP (gly-rich loop open), whilst the other appears to be incompatible with ATP binding (gly-rich loop closed) since it would sterically clash with the ß and γ phosphate moieties. The gly-rich loop open conformation interacts with the oncogenic mutation

site E768 suggesting a possible relationship between this conformation and oncogenic activation. A method for crystallisation of the RET-KD with AMPPNP and a RET activation loop substrate peptide is also described and could be followed up in the future. I note that it has not proven possible to capture a *DFG-out* conformation of the RET-KD to enable structural characterisation of the interaction of type II inhibitors such as sorafenib with RET.

### 6.3.1 Towards testing whether the RET-KD is regulated by *trans*-inhibition

All crystal forms of the RET-KD that have been determined within the laboratory show the RET-KD adopting a head-to-tail dimeric arrangement. This includes RET-KD constructs containing the five N-terminal vector-derived residues that contribute to the αN-helix and longer constructs with entirely native juxtamembrane sequences. The head-to-tail dimer is incompatible with substrate binding and could therefore form part of the RET auto-inhibitory mechanism, perhaps in ligand-free RET dimers. Whether physiologically relevant or not this dimer is evidently the favoured RET-KD arrangement in RET-KD crystallisation. Perhaps because formation of this compact dimer traps a single RET-KD activation loop conformation, whereas the monomeric forms of RET-KD may show greater flexibility and sample a number of activation loop conformers, leading to the dimer contributing to an ordered crystal lattice comparably more readily.

During this thesis I produced two double mutants of RET-KD, one of which, P766R/Q910R, weakens the formation of the head-to-tail dimer. The other, P766R/P957R, likely completely prevents head-to-tail dimer formation, as judged by the inability of this mutant to crystallise under the same conditions as the wild type RET-KD protein. *In vitro* kinase activity was not altered by either of these double mutations when compared with the wild type RET-KD. It would be instructive to obtain a crystal structure of the P766R/P957R mutant, since a new crystal lattice may allow us to trap other relevant conformations of RET-KD crystallographically. In the absence of this however, crystallisation experiments have at least indicated the impact of these mutations in perturbing the head-to-tail dimer. Informed by this, these mutations could

be useful tools to test the relevance of the crystallographic dimer on RET activity in cells.

### 6.3.2 Crystal structures of RET with *DFG-out* inhibitors

Active arrangements of protein kinase structures are highly similar due to the constraints imposed by their shared catalytic mechanism. In contrast, the inactive structures of protein kinase are highly variable and there are many distinct and unique regulatory mechanisms that operate to inhibit catalytic activity (Noble *et al.*, 2004). All crystal forms of RET-KD have an active *DFG-in* conformation and this feature may relate to and indeed be coupled with the observed head-to-tail dimer. This arrangement tethers the αC-helix into an active conformation through engagement of the αN helix with a hydrophobic patch on the αC-helix. Thus our inability to crystallise known *DFG-out* compounds such as sorafenib is perhaps not surprising.

A new RET-KD crystal form I have been able to obtain contains the RET-KD bound to an activation-loop substrate peptide. It is possible that this structure may still capture an active conformation of RET-KD, but substrate binding should preclude the head-to-tail dimer configuration thus allowing us to finally see a different RET-KD arrangement in the crystal. This should therefore be followed up by further optimisation of these crystals.

## 6.4 Future directions

### 6.4.1 Towards an atomic resolution structure of the entire RET-ECD

The SAXS-derived zRET-ECD model reveals a 100° angle between CLD3 and CLD4, however the structural determinants of this unusual cadherin arrangement are unknown. Furthermore in the context of a potential RET-CLD1:GFRα-D3 interaction in the RET ternary complex, which appears to be conserved between the mTC and zTC, an atomic resolution description of a lower vertebrate CLD1 interface for comparison with the human CLD1 interface is of significant interest. This is because it may shed light on the apparently conserved interaction despite the involvement of divergent structural elements between mammalian and lower vertebrate RET. For these reasons optimising

the zRET-CLD(1–4)$_{redsug}$ crystals presented in this thesis to allow structure determination is an important first step moving forward. Should classical crystal optimisation methods including varying the crystallisation condition, protein concentration, testing additives to the crystallisation condition and seeding not yield crystals with improved diffraction, a further strategy could be to remove more N-linked glycosylation sites from the zRET-CLD(1–4) construct. In particular the three N-glycosylation sites in CLD(1–2) are conserved between human and zebrafish and two (N98 and N199, human numbering) were shown to not be essential for hRET-CLD(1–2) production (Kjaer *et al.*, 2010). These residues may therefore be good future targets for mutation since they may not be required for zRET-ECD production, although their mutation could still negatively impact overall protein yields.

The structure of the CRD remains undetermined and this is an outstanding objective for future work. A CRD structure would be particularly informative in understanding the arrangement of this domain in the mTC and defining any potential for allosteric changes arising from ligand/co-receptor binding by fitting the structure into the EM map. It would also be a key tool in understanding the Hirschsprung's and oncogenic mutations clustered in this domain. The structure of this domain in the context of the whole ECD would be most desirable, given the apparent influence of CLD3 on the CRD arrangement and its interaction with CLD4, but even in isolation or with CLD4 the crystal structure would be informative. Towards this goal further crystallisation screening of the zRET-CLD4CRD constructs presented in this thesis may be fruitful. To pursue crystallisation of the entire zRET-ECD further, introducing the N-linked glycosylation mutations in the zRET-CLD(1–4)$_{redsug}$ construct to the zRET-ECD construct could increase the chances of crystallisation.

### 6.4.2 Defining the arrangement of RET-ECD/ligand/co-receptor in the mTC and towards atomic resolution of the complex

The immediate goals of further work on the architecture of the mTC are: to improve the resolution of the EM 3D reconstruction to allow components to be fitted in the map with greater accuracy; to define the hand of the map; and to validate an interpretation of the arrangement of the mTC components in the complex. Recent mTC preparations by Dr. Svend Kjær show improved final yields of the complex allowing an additional

purification step, which has led to improved sample homogeneity. It is hoped that this sample will allow us to improve the resolution of the 3D reconstruction by increasing the number of particle images used in the current negative-stain EM 3D reconstruction, which may improve resolution or shed light on any conformational variability limiting resolution. We could also potentially move to cryo-EM with this significantly improved sample, where the obtainable resolution is much higher than negative-stain EM (van Heel *et al.*, 2000).

Following assignment of the hand of the EM map and perhaps improvement in the resolution of the map, validation of the interpretation of the 3D reconstruction and the identified interaction sites needs to be carried out. Adding a protein tag to the C-terminus of hRET-ECD should still allow formation of the mTC and EM analysis of this tagged complex would confirm the location of the C-terminus of the CRD. The C-tail of GFRα1 is extended and therefore tagging of this component would probably not be particularly informative. Alternatively, Fab fragments (fragment antigen binding) from anti-hRET-ECD antibodies with known epitopes could be used as tags. EM analysis of a complex of the mTC with a bound Fab fragment could provide confirmatory evidence for the location of RET subdomains within the 3D reconstruction. Dr. Svend Kjær has identified several such anti-hRET-ECD antibodies well suited to forming such complexes and this will be an important future objective. In this chapter I have identified potential mTC contact sites between RET-CRD and GDNF and between RET-CLD1 and GFRα1-D3. Should the current interpretation of the mTC 3D reconstruction hold, mutagenesis studies of these interactions either using a binding assay and/or a cell-based RET activity/autophosphorylation assay would help validate the relevance of these predicted interactions. Such a cell-based assay for full-length RET has recently been established in the Structural Biology Laboratory and going forward this assay could be used to assess the functional impact of mutating the conserved residues detailed in this chapter at the proposed RET:GFRα1/GDNF contact sites.

The long-term goal in structural characterisation of a RET ternary complex is still to obtain a crystal structure of the complex, since this would provide the most detailed description of the interactions between the components in the complex. In this thesis I

have presented a method for the production of the zTC at sufficient yields for crystallisation screening using the well-established Baculovirus system with an insect cell host, which is amenable to large-scale cell production (Jarvis *et al.*, 2009). However in order to pursue crystallisation of this complex the stability will need to be improved. This may simply require buffer screening to identify a buffer in which the complex is stable, which could be carried out using a thermal stability assay (Kopec and Schneider, 2011).

Beyond this an integrated understanding of the RET receptor tyrosine kinase as a whole, combining the RET-ECD and RET-ICD work described in this thesis and by others into a single unified model of RET activation would be the target objective. This will likely require the production of the full-length RET transmembrane protein purified in detergent micelles to mimic the cell membrane environment. If a homogenous sample can be produced, this could potentially be well suited to analysis by EM in the presence or absence of soluble ligand/coreceptor complexes. Methods for production of such ligand/coreceptor complexes from both human/rat (work by Dr. Svend Kjaer) and zebrafish species have been described in this thesis. This approach, as recently reported for the EGFR (Mi *et al.*, 2011), could allow visualisation of the allosteric changes relevant to activation of both wild type and oncogenic forms of RET in the full length receptor. Alongside further high resolution structural analysis of RET subdomains and complexes this analysis could significantly develop our understanding of this intriguing receptor.

## 6.5 Concluding remarks

In this thesis I have sought to develop our understanding of the structure and regulation of the RET receptor tyrosine kinase. I have provided the first look at the overall architecture of the entire RET-ECD and the RET-ECD/ligand/co-receptor ternary complex. These data will inform future studies looking at the activation of RET by ligand/co-receptor. I have also presented the crystal structure of the RET-KD with the clinically-relevant RET inhibitor sunitinib, revealing the details of the interaction of this multi-kinase inhibitor with the RET-KD.

# Chapter 7. Materials and methods

## 7.1 Commonly used materials

Unless otherwise stated all fine chemicals and oligonucleotide primers were from Sigma-Aldrich. All restriction enzymes were from New England Biolabs (NEB).

## 7.2 Cloning

| Buffer | Composition |
|---|---|
| LB medium | 5 g Bacto yeast extract, 10 g Bacto-tryptone, 10 g NaCl, volume adjusted to 1 l with water, autoclaved |
| LB-agar | 6 g agar, 400 ml LB medium, autoclaved |
| SOC medium | 20 g Bacto-tryptone, 5 g Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM D-glucose, volume adjusted to 1 l with water, autoclaved |
| DNA 6× sample buffer | 0.25% bromophenol blue, 30% (v/v) glycerol |
| TAE buffer | 40 mM Tris-acetate, 0.1 mM EDTA |

**Table 7.1 Buffers used in cloning**

### 7.2.1 Baculovirus transfer vectors

Derivatives of baculovirus transfer vectors pBacPAK-His3 (Clontech) and pVL1393 (AB Vector) were used in this thesis. For expression of the extracellular portions of zRET and zGFRα1a and their subdomains (detailed in Table 7.2), the cDNA was cloned into pBacPAK-LL, a derivative of pBacPAK-His3 that I produced based on the expression cassette used in the mammalian cell production of hRET-CLD(1–2) (Kjaer *et al.*, 2010). The native signal sequences of zRET and zGFRα1a were used to ensure proteins were inserted into the ER and secreted efficiently following expression. At the C-terminus of the zRET-ECD or zGFRα1a constructs the pBacPAK-LL vector encodes a 3C-protease recognition site and a TEV protease recognition site separated by three glycine residues. These are followed by a Protein A tag. Initially the construct only contained a TEV protease site, however TEV cleavage of the expressed fusion proteins was ineffective. On the assumption that this was due to steric occlusion of the TEV

protease site, the linker was extended and a 3C cleavage site was added. The expressed fusion proteins could then be efficiently cleaved with TEV or 3C-protease. Cleavage of the fusion proteins produced by this expression vector with 3C-protease leaves the extraneous residues GSEFLEVLFQ at the C-terminal end; and cleavage with TEV protease leaves GSEFLEVLFQGPGGGENLYFQ. The 3C-protease was used exclusively in this thesis.

zGDNF was expressed using a derivative of pVL1393 (AB Vector) obtained from Dr. Svend Kjær, which is referred to as pBac, as its baculovirus transfer vector. This construct encodes a honeybee melittin signal sequence, which is an effective secretion signal sequence for proteins expressed in insect cells (Tessier *et al.*, 1991). This construct was used for zGDNF expression since zGDNF is natively expressed as a proprotein and subcloning of the mature protein gene sequence would remove the native signal sequence as well as the pro-sequence.

The intracellular hRET-KD was cloned into a pBacPAK-His3 derivative referred to as pBacPAK-GST. This vector was obtained from Mr. Phillip Knowles and was used previously in the production of hRET-KD in the Structural Biology Laboratory (Knowles *et al.*, 2006). However in order to improve the expression levels of hRET-KD, a codon optimised gene of hRET-KD for insect cell expression was obtained from GeneART and freshly subcloned into this vector. The pBacPAK-GST encodes an N-terminal glutathione S-transferase (GST) tag followed by a 3C cleavage site. Following cleavage of the fusion protein with 3C-protease the extraneous residues GPLSL are left at the N-terminus.

### 7.2.2 cDNA clones

The zRET cDNA was a kind gift of Dr. Tiffany Heanue and Dr. Vassilis Pachnis (National Institute of Medical Research). The zGFR$\alpha$1a and zGDNF cDNAs were a kind gift of Dr. Iain Shepherd (Emory University, US). zRET-CLD(3–4) containing the N-linked glycosylation site mutations N259Q/N308Q/N390Q/N433Q was purchased from GeneART and was codon-optimised for insect cell host expression. This was subcloned into the zRET-CLD(1–4) construct to produce the zRET-CLD(1–4)$_{\text{redsug}}$

construct. The hRET-ICD was also codon-optimised for insect cell expression and was purchased from GeneART.

| Construct name | Expression cassette details |
|---|---|
| pBacPAK-LL-zRET-CLD(1–2)$_{1–249}$ | zRET$_{1–249}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-CLD(1–2)$_{1–253}$ | zRET$_{1–253}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-CLD(1–4) | zRET$_{1–504}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-CLD(1–4)$_{redsug}$ | zRET$_{1–504}$ N259Q/N308Q/N390Q/N433Q-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-ECD | zRET$_{1–626}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-CLD4CRD$_{373–626}$ | zRET$_{373–626}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zRET-CLD4CRD$_{382–626}$ | zRET$_{382–626}$-*3C*-*TEV*-protein A |
| pBac-zGDNF | Melittin SS-zGDNF$_{89–235}$ |
| pBac-zGDNF$_{\Delta N}$ | Melittin SS-zGDNF$_{135–235}$ |
| pBacPAK-LL-zGFRα1a$_{\Delta C}$ | zGFRα1a$_{1–352}$-*3C*-*TEV*-protein A |
| pBacPAK-LL-zGFRα1a$_{D2D3}$ | zGFRα1a$_{1–21;143–352}$-*3C*-*TEV*-protein A |
| pBacPAK-GST-RET-KD | GST-*3C*-hRET$_{705–826;841–1013}$ (codon optimised) |
| pBacPAK-GST-RET-KD$_{P766R/Q910R}$ | GST-*3C*-hRET$_{705–826;841–1013}$ P766R/Q910R (codon optimised) |
| pBacPAK-GST-RET-KD$_{P766R/P957R}$ | GST-*3C*-hRET$_{705–826;841–1013}$ P766R/P957R (codon optimised) |

**Table 7.2 Baculovirus transfer vector constructs**

pBacPAK-LL and pBacPAK-GST = modified pBacPAK-His3 (Clontech) vectors; pBac = modified pVL1393 vector (AB Vector); *3C* = 3C-protease recognition site; *TEV* = TEV protease recognition site; GST = glutathione S-transferase; Melittin SS = honeybee melittin signal sequence.

### 7.2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) amplification of a target gene for cloning into a different plasmid vector was carried out in a thermocycler (Peltier). Typically 50 µl reaction mixtures contained 1× Thermopol buffer (NEB), 0.4 µM forward primer, 0.4 µM reverse primer, 2 µM deoxyribonucleic triphosphates (dNTPs), 2 units VENT™ polymerase (NEB) and 100 ng template DNA. PCR was carried out with an initial denaturation at 94 °C for 1 minute and annealing at 5 °C below the primer melting temperature for 1 minute, followed by 30 cycles of the following sequence: extension at 72 °C for 1 minute per 1 kb of target gene length, denaturation at 94 °C for 30 seconds and annealing at 5 °C below the primer melting temperature for 30 seconds. This was followed by a final extension step at 72 °C for 10 minutes and storage at 4 °C. Template

DNA was then removed from the reaction mix by digestion with 20 units of Dpn1 for at least 4 hours.

### 7.2.4 Cloning into baculovirus transfer vectors

PCR products of the target gene, and the target baculovirus transfer vector, were digested with restriction enzymes for ligation. Nco1 and BamH1 enzymes were used for cloning into the pBacPAK-LL vector; HindIII and XhoI for the pBacPAK-GST vector; and XbaI and Not1 for the pBac vector. The target vector DNA was prepared using the Qiagen maxi prep kit. Digestion reactions contained 1× NEB reaction buffer, 20 units of each enzyme, 1 mg/ml bovine serum albumin and either 45 µl of the PCR reaction mix or 50 µl of the desired vector (typically 50 µg DNA) to a final volume of 60 µl. Digestion reactions were incubated for at least 4 hours at 37 °C. 10 units of calf intestinal alkaline phosphatase was then added to the digested vector and allowed to dephosphorylate the cut ends of the vector for one hour at 37 °C. Digestion products of the target gene and target vector were purified by gel electrophoresis. 1× DNA loading buffer was added to the digest reactions and a DNA ladder, and these were run on a 0.8% agarose gel containing ethidium bromide in TAE buffer at 120 V for 30–40 minutes. Bands corresponding to the digested gene and vector were cut out of the gel and the agarose removed using Ultrafree DA columns (Millipore). The DNA was further purified from the gel extracts using the final steps of the Qiagen mini-prep kit protocol: four-fold excess of buffer PB was added to the gel extract; this was applied to a mini-prep DNA binding spin column and spun for 1 minute at 1000 × g and 1 minute at 16000 × g; the column was then washed with 750 µl buffer PE; digested DNA was then eluted in 30 µl buffer EB for the vector and 20 µl for the gene. DNA concentration of the purified digested DNA was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific), based on absorbance at 260 nm.

The purified digested vector and gene were then ligated together. The 20 µl ligation reactions contained a 1:6 molar ratio of vector:gene, 1× T4 ligation buffer (NEB) and 400 cohesive end units of T4 ligase (NEB). A control reaction without gene was also carried out in tandem to determine the background level of undigested vector present. These were incubated at 16 °C for 3 hours followed by transformation into 250 µl of chemically competent NovaBlue cells. Transformation was carried out by a 45 second

incubation at 42 °C followed by 2 minutes on ice. The transformed cells were then amplified by addition of 500 µl SOC medium and incubation at 37 °C for 1 hour, followed by spreading on LB-agar plates with 200 ug/ml carbenicillin (for pBacPAK constructs) and incubation at 37 °C overnight.

If ligation was successful there should be many more colonies on the ligation reaction plates compared to the control plate, with the control plate only containing a handful or no colonies. If this was the case, plasmids from 10 colonies on the ligation reaction plate were typically amplified using the Qiagen mini-prep kit following manufacturers instructions. These were then assessed to see whether the ligation had been successful by DNA sequencing (section 7.2.7).

### 7.2.5 Site-directed mutagenesis

Site-directed mutagenesis of plasmids, including insertions or deletions of up to 33 bases pairs, was carried out by QuikChange™ (Stratagene). Typically reaction mixtures contained 1× PFU TURBO buffer (Stratagene), 0.4 µM forward primer, 0.4 µM reverse primer, 2 µM deoxyribonucleic triphosphates (dNTPs), 2.5 units PFU TURBO polymerase (Stratagene), 100 ng template DNA and 2 mM $MgSO_4$. The QuikChange™ PCR reactions were carried out with an initial denaturation at 96 °C for 2 minutes and annealing at 55 °C for 2 minutes, followed by 17 cycles of the following sequence: extension at 68 °C for 1 minute per 1 kb of template DNA plus 1 minute, denaturation at 96 °C for 1 minute and annealing at 55 °C for 1 minute. This was followed by a final extension step at 68 °C for 20 minutes and storage at 4 °C. Template DNA was then removed from the reaction mix by digestion with 20 units of Dpn1 for at least 4 hours at 37 °C. 250 µl of chemically competent NovaBlue cells were then transformed with 5 µl of the reaction mix by incubation at 42 °C for 45 seconds followed by 2 minutes on ice. Transformed cells were spread on LB-agar plates containing 200 ug/ml carbenicillin (for pBacPAK constructs) and incubated at 37 °C overnight. Plasmids were purified using the Qiagen mini-prep kit according to the manufacturers instructions and sequenced (section 7.2.7) to determine whether mutagenesis had been successful.

### 7.2.6 Large deletions from plasmid constructs

To eliminate an intron found in the zRET cDNA a modified QuikChange™ (Stratagene) protocol was used, allowing deletion of large portions of DNA. The primers used did not span the deletion site, as in standard QuikChange™, but were designed for amplification of a linear product containing the entire plasmid and gene sequence except the sequence to be deleted. The same reaction mix and thermocycler sequence as for section 7.2.5 were used for the PCR reaction, which was followed by Dpn1 digestion of the template with 20 units of Dpn1. To recyclise the linear PCR product, the ends were firstly phosphorylated by addition of 1× T4 ligase buffer (NEB) and 10 units of T4 polynucleotide kinase to the PCR reaction mix for 30 minutes at 37 °C. Ligation was then carried out using 2–6 µl reaction mix in 1× T4 ligase buffer (NEB) with 1µl T4 ligase at 16 °C for 3 hours with a no ligase control. Ligation products were transformed into NovaBlue cells and assessed as described in section 7.2.4.

### 7.2.7 DNA sequencing

The open reading frames of all constructs were verified by DNA sequencing. The Cancer Research UK London Research Institute equipment park staff carried out all DNA sequencing, using the BigDye™ Terminator Cycle Sequencing kit (PE Applied Biosystems) and capillary sequencing on an ABI Prism 3730.

## 7.3 Baculovirus method for insect cell expression

### 7.3.1 Baculovirus production

Production of baculoviruses from the baculovirus transfer vector constructs, initial amplification of baculoviruses and protein expression testing in insect cells were carried out by staff at the Cancer Research UK London Research Institute Protein Production Facility. Baculoviruses were produced using the Clontech BacPAK system in Sf9 cells (BacPAK™ Expression System User Manual, Clontech).

### 7.3.2 Insect cell culture

Sf9 cells were cultured in suspension in serum-free SfIII media (Invitrogen) supplemented with 10 μg/ml gentamycin. They were grown at 27 °C with shaking (140 rpm) in 100–500 ml aliquots in 2 l roller bottles. The cells were kept at a cell density of between 1.5–6 × $10^6$ cells/ml. Hi5 cells were cultured in suspension in the same way as the Sf9 cells, but they were grown in SfII media (Invitrogen) and only switched to SfIII media for protein expression. This was because cell doubling rates and viability was poor when the Hi5 cells were cultured in SfIII media for long periods, but the yields from protein expression by Baculovirus infection were increased by 50–100% when carried out in SfIII media. Cell density and viability was determined using a Vi-CELL (Beckman Coulter). Cells that were observed to have an approximate doubling time of less than 24 hours or a viability of less than 89 % were not used in baculovirus amplification or protein expression.

### 7.3.3 Baculovirus amplification

Baculoviruses were amplified in Sf9 cells. The baculovirus was added at a MOI of 0.1–0.5 to Sf9 cells at 1 × $10^6$ cells/ml in 250 ml aliquots. The culture was incubated for 6 days at 27 °C with shaking (140 rpm). The baculovirus containing media was then harvested by centrifugation (3000 × g, 20 minutes, 4 °C). This amplification process was often repeated: Baculoviruses used in protein expression had been passaged in Sf9 cells between 3 and 5 times.

### 7.3.4 Calculation of viral titre

The High Pure Viral Nucleic acid Large Volume Kit (Roche) was used to purify baculoviruses from serum following manufacturers instructions. RT-PCR was then used to determine viral titre following manufacturers instructions.

### 7.3.5 Protein expression

Hi5 cells at 1 × $10^6$ cells/ml in 500 ml aliquots were infected with baculoviruses at a MOI of 2 for all proteins except zGFRα1a$_{\Delta C}$/zGDNF$_{\Delta N}$ co-infection. In order to achieve similar levels of expression for zGFRα1a$_{\Delta C}$ and zGDNF$_{\Delta N}$ when co-expressed,

zGFRα1a$_{\Delta C}$ baculovirus was added with an MOI of 1 and zGDNF$_{\Delta N}$ baculovirus with an MOI of 5. The infected cells were incubated at 27 °C with shaking (140 rpm) for 3.5 days for intracellular proteins and for 4.5 days for extracellular proteins. The protein expression yields following affinity purification (described below) were: 4 mg/l culture for zRET-ECD and zRET-CLD(1–4); 1 mg/l culture for zRET-CLD(1–4)$_{redsug}$; 1.4 mg/l for both zRET-CLD4CRD constructs; 1 mg/l culture zGFRα1a/zGDNF coexpression; 4 mg/l culture hRET-KD wild type and mutants.

## 7.4 Protein purification

### 7.4.1 Buffers

| Name | Composition |
|---|---|
| **ECD buffer** | 20 mM Tris pH 7, 200 mM NaCl, (±2 mM $CaCl_2$) |
| **Anion exchange buffer A** | 20 mM Tris pH 7, 50 mM NaCl, 5% (v/v) glycerol, (±2 mM $CaCl_2$) |
| **Anion exchange buffer B** | 20 mM Tris pH 7, 500 mM NaCl, 5% (v/v) glycerol, (±2 mM $CaCl_2$) |
| **KD lysis buffer** | 50 mM Tris pH 8, 100 mM NaCl, 1 mM dithiothreitol (DTT), 10 mM Benzamidine, 0.2 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) |
| **KD buffer** | 20 mM Tris pH 8, 100 mM NaCl, 1 mM DTT |
| **mTC buffer** | 25 mM Tris, pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$ |
| **GST-Endo-F1 lysis buffer** | 50 mM Tris pH 7.5, 1 × Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail tablet (Roche) |
| **GST-Endo-F1 size exclusion buffer** | 50 mM Tris pH 7.5, 200 mM NaCl |
| **GST-Endo-F1 storage buffer** | 10 mM Sodium acetate pH 5.5, 50 % glycerol |

**Table 7.3 Protein purification buffer compositions**

### 7.4.2 Extracellular protein purification

The following method was used for purification of the zRET-ECD, CLD(1–2), CLD(1–4) and CLD4–CRD and zGFRα1a (± untagged zGDNF) Protein A tagged proteins. For purification of zRET-ECD and CLD(1–4) the buffers contained 2 mM $CaCl_2$.

#### *7.4.2.1 Protein A affinity purification*

Following baculovirus infection the media was harvested by centrifugation to remove the cells (3000 × g, 20 min, 4 °C). The clarified media was incubated in 2 l batches divided into 250 ml aliquots with IgG-sepharose resin (GE Healthcare). 1 ml resin was used per litre of media and each 2 l batch was incubated at 4 °C with rotation for 1.5–2 hours. The resin was then gently spun down (250 rpm, 10 min, 4 °C), the media removed and fresh media added. This was repeated until all the media had been incubated with the resin. The resin was then placed in a 10 ml syringe with a 0.2 μm filter on ice and washed with ice-cold ECD buffer (5 × 10 ml). The resin was resuspended in 10 ml ECD buffer and transferred to a 15 ml falcon. GST-tagged human rhinovirus 3C-protease (PreScission protease, Amersham Biosciences) was added, 100 μg per 1 ml resin, and allowed to cleave overnight at 4 °C with rotation.

Cleaved material was separated from the resin in a 10 ml syringe, and the resin washed with at least 2 column volumes of ECD buffer. To remove the GST-tagged protease, 200 μl of ECD buffer washed glutathione-sepharose resin (GE Healthcare) was added to the eluate and incubated for 45 minutes at 4 °C with rotation. Unbound material was eluted from the resin in ECD buffer using a 10 ml syringe with a 0.2 μm filter. This affinity purified material was then concentrated using Vivaspin centrifugation columns and either subjected to further purification by size exclusion chromatography or used directly in binding experiments.

#### *7.4.2.2 Size exclusion chromatography*

Affinity purified protein was concentrated to 5 ml for size exclusion on a 150 ml Superdex 200 26/300 column or 0.25 ml for a 24 ml Superdex 200 10/300 or Superose 6 10/300 GL column (GE Healthcare). Size exclusion was carried out at 4 °C. The column was pre-equilibrated with ECD buffer, the sample was then injected and buffer flowed at 2 ml/minute for the large column and 0.4 ml/minute for the smaller column, using the ÄKTA purifier system (GE Healthcare). 2 ml or 0.4 ml fractions were collected. Protein elution was followed using by absorbance at 280 nm (Abs280) measurement. Peak fractions were combined and concentrated for anion exchange chromatography or used in further experiments directly.

#### *7.4.2.3 Anion exchange chromatography*

Protein was concentrated to a volume of 250–500 µl and diluted in 5–10 ml anion exchange buffer A. A 1 ml HiTrap Q HP (GE Healthcare) at 4 °C was equilibrated with anion exchange buffer A. The protein was loaded onto the column at a flow rate of 0.4 ml/minute and unbound protein washed off with anion exchange buffer A. Bound protein was then eluted by applying a gradient of 100% anion exchange buffer A to 100% anion exchange buffer B over 20 column volumes, collecting 0.4 ml fractions. Protein elution was followed using the Abs280 measurement. Peak fractions were combined and concentrated for use in further experiments.

#### *7.4.2.4 Deglycosylation with endoglycosidase F1*

Endoglycosidase F1 (Endo-F1) deglycosylation of zRET-ECD proteins was carried out using GST-Endo-F1 produced in bacteria using standard protocols from a construct kindly donated by Prof. Yvonne Jones (see section 7.4.5). Following affinity purification (section 7.4.2.1), GST-Endo-F1 was added to the zRET protein at a ratio of 1:5 (weight:weight) in ECD buffer and incubated for 20 hours at 4 °C with rotation. The GST-Endo-F1 was then removed by addition of 100–200 µl of glutathione-sepharose resin and incubation for 30 minutes at 4 °C with rotation; followed by separation of the supernatant from the resin using a 10 ml syringe with a 0.2 µm filter. The supernatant containing the enzymatically partially-deglycosylated zRET protein was then subject to further purification by size exclusion chromatography (section 7.4.2.2).

### 7.4.3 Zebrafish ternary complex assembly and purification

For standard large-scale production of the zTC (zebrafish ternary complex = zRET-ECD/zGFRα1a$_{\Delta C}$/zGDNF$_{\Delta N}$) 5.5 l of zGFRα1a$_{\Delta C}$-Protein A/zGDNF$_{\Delta N}$ and 2 l of zRET-ECD-Protein A Baculovirus infected cell culture was treated as detailed in section 7.4.2.1: The media was clarified by centrifugation and Protein A fusion proteins were separately affinity-purified at 4 °C using IgG-sepharose resin (5 ml for GFRα1a$_{\Delta C}$/GDNF$_{\Delta N}$ and 2 ml for zRET-ECD). Typically the zRET-ECD was purified in this manner the day before the zGFRα1a$_{\Delta C}$/zGDNF$_{\Delta N}$. The zRET-ECD-Protein A bound resin was then washed with ice-cold ECD buffer (5 × 10 ml) and zRET-ECD

eluted in 10 ml ECD buffer by addition of 3C-protease overnight as in section 7.4.2.1. The zGFRα1a$_{\Delta C}$-Protein A/zGDNF$_{\Delta N}$ bound resin was similarly washed with ice-cold ECD buffer (5 × 10 ml), but prior to addition of 3C-protease, the eluted zRET-ECD in 10–15 ml ECD buffer was added to the washed IgG-sepharose resin with zGFRα1a$_{\Delta C}$-Protein A/zGDNF$_{\Delta N}$ bound in a 50 ml falcon. zRET-ECD was allowed to bind for one hour with rotation at 4 °C. The resin was then placed in a 10 ml syringe on ice and the unbound material was eluted. The resin was washed with at least five column volumes of ECD buffer, and resuspended in 2–3 column volumes of ECD buffer. As per section 7.4.2.1–7.4.2.3, 3C-protease was then added to the zGFRα1a$_{\Delta C}$-Protein A/zGDNF$_{\Delta N}$/zRET-ECD resin to elute the complex and purification continued with size exclusion chromatography.

### 7.4.4 hRET-KD protein purification

The protocol used for production of GST tagged hRET-KD wild type and mutant proteins and is essentially the same as the previously published hRET-KD purification protocol (Knowles *et al.*, 2006).

#### *7.4.4.1 Cell lysis and GST affinity purification*

Cells were harvested following baculovirus infection by centrifugation (2000 × g, 12 min, 4 °C). Cell pellets from 1 l cell culture were resuspended in 20 ml ice cold KD lysis buffer and lysed by sonication for 40–60 seconds on ice. The cell lysate was clarified by centrifugation (29220 × g, 30 min, 4 °C).

0.5 ml glutathione-sepharose resin (GE Healthcare) per l original cell culture was washed with KD lysis buffer and added to the clarified lysate. This mixture was incubated at 4 °C with rotation for 1.5 hours. The resin was then collected in a 10 ml syringe and washed with KD buffer (5 × 10 ml). The resin was subsequently resuspended in KD buffer and GST tagged 3C-protease 100 μg per 1 ml resin was added. Cleavage was allowed to occur overnight at 4 °C with rotation. Cleaved protein was then separated from the resin using a 10 ml syringe, and the resin washed with at least 2 column volumes of KD buffer.

#### *7.4.4.2 Autophosphorylation and dialysis*

The affinity-purified hRET-KD was phosphorylated by addition of 5 mM ATP (pH 8) and 10 mM $MgCl_2$ and incubated on ice for 6 hours. The ATP and magnesium were then removed by dialysis of the protein solution against 1 l KD buffer overnight in a 12.5 kDa molecular weight cut off dialysis membrane at 4 °C.

Non-phosphorylated hRET-KD was produced by incubation with calf intestinal alkaline phosphatase (NEB) for 5 hours at 4 °C whilst protein was still bound to the glutathione-sepharose resin. The enzyme was then washed off with 2 × 10 ml KD buffer and 3C cleavage carried out as in section 7.4.4.1.

#### *7.4.4.3 Size exclusion chromatography*

Size exclusion chromatography of hRET-KD was carried out as for the extracellular proteins (section 7.4.2.2) but using Superdex 75 columns (GE Healthcare) and running the column in KD buffer with 400 mM NaCl for the RET-$KD_{P766R/P957R}$ mutant.

### 7.4.5 GST-endoglycosidase F1 production

The pGEX endoglycosidase F1 bacterial expression construct with pRep4 was kindly donated by Prof. Yvonne Jones (Oxford University), originally produced by Dr. Yoav Peleg (Weismann Institute, Israel). The received plasmids were transformed into NovaBlue cells (as section 7.2.4) and the transformed cells were spread on an LB-agar plate with 50 μg/ml kanamycin and 100 μg/ml ampicillin, to select for colonies containing both plasmids, and incubated at 37 °C overnight. A single colony was amplified overnight in 100 ml LB with 50 μg/ml kanamycin and 100 μg/ml ampicillin and purified by maxi-prep (Qiagen) following manufacturers instructions.

#### *7.4.5.1 Protein expression in bacterial cells*

GST-endoglycosidase F1 (GST-Endo-F1) expression and purification was carried out based on Grueninger-Leitch *et al.*, 1996. FB810 cells (a RecA- strain of BL21) were transformed with maxi-prep prepared plasmids pGEX endoglycosidase F1 and pRep4 by electroporation in a chilled cuvette. 0.5 ml of SOC media was then added and cells were incubated in an eppendorf at 37 °C for 1 hour. 50 μl of the cell culture with 50 μl

SOC media were spread on an LB-agar plate with 50 μg/ml kanamycin and 100 μg/ml ampicillin and incubated at 37 °C overnight. A single colony from transformation was amplified overnight in 100 ml LB with 50 μg/ml kanamycin and 100 μg/ml ampicillin. 40 ml of this culture was then used to inoculate 2 l of LB containing 50 μg/ml kanamycin and 100 μg/ml ampicillin for protein expression. The 2 l culture was grown in 1 l aliquots in 2 l flasks at 37 °C until the optical density at 550 nm (OD550) reached 0.8. Protein expression was then induced by addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells were incubated for 4 hours at 37 °C. Cells were harvested by centrifugation (4000 × g, 20 minutes, 4 °C).

#### *7.4.5.2 Cell lysis and protein purification*

GST-Endo-F1 cells were resuspended in 20 ml ice-cold GST-Endo-F1 lysis buffer per 1 l original cell culture and lysed by sonication on ice (2 minutes, 15 microns). The cell lysate was clarified by centrifugation (29220 × g, 30 min, 4 °C). 8 ml of glutathione-sepharose resin (GE Healthcare) washed with GST-Endo-F1 lysis buffer was added to the clarified lysate in a 50 ml falcon and this was incubated at 4 °C for 1 hour with rotation. The resin was spun down (500 × g, 5 min, 4 °C) and the supernatant discarded. The resin was then washed with 3 × 50 ml GST-Endo-F1 lysis buffer. The GST-Endo-F1 was eluted from the resin by incubation in 2 × 30 ml GST-Endo-F1 lysis buffer with glutathione for 30 minutes at 4 °C with rotation, and the supernatant collected by centrifugation (500 × g, 5 min, 4 °C).

The eluted GST-Endo-F1 was concentrated in a Vivaspin column to 5 ml and loaded on a Superdex 75 26/300 column (GE Healthcare) pre-equilibrated with GST-Endo-F1 size exclusion buffer. Size exclusion chromatography was carried out using an ÄKTA purifier system (GE Healthcare) with a flow rate of 2 ml/minute and 2 ml fractions were collected. Fractions from the GST-Endo-F1 peak (as confirmed by SDS-PAGE) were combined and dialysed against 1 l GST-Endo-F1 storage buffer overnight at 4 °C.

The final yield of GST-Endo-F1 following dialysis was 36 mg/l culture. The protein was divided into 100 μl aliquots at 4.8 mg/ml, frozen on dry ice and stored at -80 °C.

## 7.5 Mammalian extracellular protein production

Dr. Svend Kjær produced the human RET-ECD and mammalian ternary complex (mTC) used in this thesis. His methods are described below.

### 7.5.1 Mammalian cell protein expression

The hRET-ECD construct used in this thesis contained the cysteine mutations C87R and C216R. These mutations remove the two unpaired cysteine residues from the hRET-ECD and therefore improve expression levels by reducing the tendency of hRET-ECD to misfold. These mutations do not affect GFRα1/GDNF binding affinity (Kjaer *et al.*, 2010). The rGFRα1$_{1–425}$ construct contains the entire rGFRα1 sequence except the C-terminal GPI-link site and GPI signal sequence. The rGFRα1$_{1–425}$ construct contains the mutation N59Q, which removes a partially occupied N-linked glycosylation site.

hRET-ECD C87R C216R (Kjaer *et al.*, 2010) and rGFRα1$_{1–425}$ N59Q were expressed as TEV cleavable Protein A fusion proteins in stably transfected Chinese Hamster Ovary (CHO) Lec8 cells (Stanley *et al.*, 1991). The stable cell lines were generated by transfecting CHO cells with the pcDNA3 constructs (Table 7.4) using the Effectene reagent (Qiagen), followed by selection with 500 μg/ml G418 (Invitrogen). For protein production the cells were grown in 2 l roller bottles for 4–5 days in standard DMEM, 10% (v/v) FBS, 50 μg/ml gentamycin and the supernatant was subsequently harvested by centrifugation and filtered with 0.22/0.45 μm filters.

The mature hGDNF corresponding to residues 77–211 (plus an additional N-terminal methionine residue) was kindly provided by Amgen (Thousand Oaks, California) produced recombinantly in *Escherichia coli*.

| Construct name | Expression cassette details |
|---|---|
| pcDNA3-hRET-ECD C87R C216R | hRET-ECD$_{1–635}$ C87R C216R-*TEV*-protein A |
| pcDNA3-rGFRα1$_{1–425}$ N59Q | rGFRα1$_{1–425}$ N59Q-*TEV*-protein A |
| pcDNA3-rGFRα1$_{\Delta D1}$ | rGFRα1$_{1–20;145–425}$-*TEV*-protein A |

**Table 7.4 Mammalian expression vectors**

TEV cleavage of the protein A tag from proteins expressed using these constructs leaves these extraneous residues at the C-terminus: EFENLYFQ.

### 7.5.2 Mammalian ternary complex assembly and purification

For a standard preparation of ternary complex, 2 l of supernatant containing either hRET-ECD-*TEV*-Protein A or rGFRα1-*TEV*-Protein A was incubated overnight at 4 °C with 0.5 ml IgG-sepharose resin (GE Healthcare). The resin was subsequently isolated by slow-speed centrifugation and washed 3 times with mTC buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$) in a 50 ml falcon. For production of the hRET-ECD, 100 µg of TEV protease was added to the resin overnight at 4 °C in 5 ml of mTC buffer and the supernatant containing hRET-ECD kept. For assembly of the mTC, washed IgG-sepharose resin with bound rGFRα1-*TEV*-Protein A was incubated with 200 µg of recombinant human GDNF (kindly provided by Amgen) for 1 hour and then washed with mTC buffer. Finally, the supernatant containing hRET-ECD was added to the rGFRα1/hGDNF resin for 1 hour, the resin was washed 3 times with mTC buffer in a 50 ml falcon and the hRET-ECD/rGFRα1/hGDNF ternary complex eluted from the resin by addition of 100 µg of TEV overnight at 4°C. The complex was subsequently purified by size exclusion chromatography in mTC buffer using a Superose 6 10/300 GL column (GE Healthcare) connected to an ÄKTA Explorer system (GE Healthcare).

### 7.5.3 Human RET-ECD purification

For SAXS analysis the hRET-ECD was produced as described above and further purified following TEV cleavage from the IgG-sepharose resin by size exclusion chromatography in mTC buffer using a Superdex 200 10/300 column (GE Healthcare).

## 7.6 Protein analysis

### 7.6.1 Measurement of protein concentrations

Protein concentrations were calculated based on the Abs280, measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific), using the protein molecular weights and extinction coefficients as detailed in Table 7.5.

| Protein | No. of amino acids | Protein MW (kDa) | Theoretical pI | Theoretical Extinction Coefficient (*$10^3$ $M^{-1}$ $cm^{-1}$) |
|---|---|---|---|---|
| zRET-ECD | 614 | 69.34 | 5.07 | 66.25 |
| zRET-CLD(1–2)$_{1–249}$ | 237 | 27.00 | 5.52 | 24.79 |
| zRET-CLD(1–2)$_{1–253}$ | 241 | 27.48 | 5.09 | 24.79 |
| zRET-CLD(1–4) | 490 | 55.65 | 5.08 | 52.43 |
| zRET-CLD(1–4)$_{redsug}$ | 490 | 55.70 | 5.08 | 52.43 |
| zRET-CLD4CRD$_{373–626}$ | 264 | 29.54 | 4.99 | 28.52 |
| zRET-CLD4CRD$_{382–626}$ | 255 | 28.51 | 4.91 | 28.52 |
| zGFRα1a$_{\Delta C}$ | 344 | 38.74 | 8.44 | 31.09 |
| zGFRα1a$_{D2D3}$ | 222 | 24.75 | 8.06 | 14.66 |
| zGDNF | 149 | 17.06 | 9.32 | 6.34 |
| zGDNF$_{\Delta N}$ | 104 | 11.92 | 7.15 | 6.34 |
| zTC (2 zGDNF$_{\Delta N}$/2 zGFRα1a$_{\Delta C}$/ 2 zRET-ECD) | 2124 | 239.90 | 5.89 | 207.48 |
| hRET-ECD | 616 | 68.84 | 5.52 | 75.55 |
| mTC (2 hGDNF/2 rGFRα1a/ 2 hRET-ECD) | 2378 | 264.34 | 7.65 | 236.53 |
| hRET-KD | 300 | 34.30 | 9.04 | 49.39 |

**Table 7.5 Protein parameters**

Theoretical biophysical parameters of expressed proteins following cleavage of the signal sequences and affinity tags. Th extinction coefficient assuming all cysteines form disulphide bonds is given for extracellular proteins and assuming all cysteines are reduced for intracellular proteins. Parameters calculated by ProtParam tool in ExPASY (http://web.expasy.org/protparam).

### 7.6.2 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using either 4–12 % NuPAGE gels (Invitrogen) with MOPs running buffer (Invitrogen) or 4–20 % NuSep gels (NuSep) with SDS running buffer. Protein samples were prepared by addition of 5× SDS loading buffer and incubation at 95 °C for 5 minutes. Samples were loaded in the wells of the gel and a voltage of 100–150 V was applied until the SDS loading buffer dye had migrated to the bottom of the gel. Gels were stained with PhastGel™ Blue R-350 Coomassie staining solution (GE Healthcare) for 15 minutes and destained with a 10% acetic acid, 40% methanol solution for 1–1.5 hours.

| Name | Composition |
|---|---|
| **5× SDS loading buffer** | 5% (w/v) SDS, 0.2 M Tris pH 6.8, 33% Glycerol, 2 mg/ml bromophenol blue (+ β-mercaptoethanol for reducing gels) |
| **10× SDS running buffer** | 121 g Tris Base, 10 g SDS, 238g HEPES, adjusted to 1 l with deionised water |
| **Coomassie staining solution** | One PhastGel™ Blue R-350 tablet (GE Healthcare) was dissolved in 80 ml water, 120 ml methanol added and solution was stored at 4 °C. Before use 50 ml of this solution was added to 50 ml 20% acetic acid. |
| **Destain solution** | 10% acetic acid, 40% methanol solution |
| **Cathode buffer** | 1× NativePAGE™ running buffer, 1× NativePAGE™ cathode buffer additive (includes 0.02 % Coomassie G-250) |
| **Anode buffer** | 1× NativePAGE™ running buffer |

**Table 7.6 SDS-PAGE and Native PAGE buffers**

### 7.6.3 Native PAGE

Native PAGE was carried out at 4 °C using the NativePAGE™ reagents from Invitrogen. Protein samples were prepared by addition of 4× NativePAGE™ sample buffer, and loaded in the wells of a 4–16 % NativePAGE™ Novex® Bis-Tris gel. Chilled Coomassie-containing cathode buffer (Table 7.6) was added to the inner buffer chamber of the gel tank and chilled anode buffer was added to outer chamber. A voltage of 150 V was applied to the gel for 105–120 minutes. The gel was then destained with destain solution.

### 7.6.4 Dynamic light scattering

DLS measurements were carried out in a Zetasizer Nano ZS (Malvern Instruments) and analysis was performed using the manufacturer-supplied software. 12 µl protein samples of > 0.5 mg/ml were placed in a quartz cuvette for measurement. This analysis provides an imperfect estimate of particle diameter and molecular weight of proteins in solution, but is particularly helpful in assessing whether a protein sample is aggregated.

### 7.6.5 Size exclusion chromatography coupled with multi-angle light scattering

SEC-MALS was used to determine the molecular weight of proteins as they elute from a size exclusion chromatography column. The SEC-MALS system used was an ÄKTA purifier system (GE Healthcare), with a Superose 6 10/300 GL column (GE Healthcare),

coupled to a Optilab®T-rEX (Wyatt Technology), which measures differential refractive index and is used to determine protein concentration, and a DAWN® HELEOS® II (Wyatt Technology) MALS unit. The system was equilibrated in ECD buffer (for analysis of zRET and zTC samples) at a flow rate of 0.1–0.5 ml/min until the differential refractive index had stabilised. 100 µl of protein sample at ~3 mg/ml was loaded onto the Superose 6 column and the SEC-MALS system was run at a 0.5 ml/min flow rate for 60 minutes. The SEC-MALS data was analysed using the manufacturer-supplied software to determine the molecular weight of the protein across the size exclusion peak.

### 7.6.6 Circular dichroism

Dr. Stephen Martin at the National Institute of Medical Research, London, carried out circular dichroism experiments. Protein samples at ~0.15 mg/ml in ECD buffer were placed in a 200 µl fused silica cuvette (Hellma, Jena) with 1 mm path length. Far-UV circular dichroism measurements were recorded on a Jasco J-715 spectropolarimeter thermostatted by a PTC 348-WI Peltier unit. Spectra were recorded with 0.2 nm resolution and the spectrum of buffer alone was subtracted from the protein spectra. Circular dichroism data processing was carried out using Dr. Stephen Martin's software (unpublished). Secondary structure prediction from circular dichroism data was carried out using SELCON, CONTIN and CDSSTR programs, which use different algorithms to compare the experimental circular dichroism spectrum to circular dichroism spectra of a number of reference proteins with known secondary structure content (Sreerama and Woody, 2000).

### 7.6.7 Ellmans assay for detecting free thiols

Mr. Phillip Knowles carried out the Ellmans assay on monomeric and dimeric zRET-CLD4CRD$_{373–626}$. The reagent mixture contained 100 mM Tris pH 8, 5.4 M Guanidium chloride, 0.5 mM Ellman's reagent (from 1 M stock dissolved in DMSO). 900 µl of the reagent mixture was added to 100 µl of protein in ECD buffer or ECD buffer alone. A positive control of 900 µl reagent mixture with 100 µl 0.25 mM reduced glutathione was also carried out. The absorbance at 412 nm (Abs412) was taken for each sample, the Abs412 of the buffer only sample was deducted, and the concentration of free thiols

calculated using the extinction coefficient of the Ellman's reagent derivative 2-nitro-5-thiobenzoate, produced in the presence of free thiols, at high salt (13700 $M^{-1}cm^{-1}$).

### 7.6.8 Zebrafish ternary complex assembly assay

The zTC assembly assay was carried out using affinity purified zRET-ECD, purified in ECD buffer with no added calcium, and affinity purified zRET-CLD(1–4) in ECD buffer with 2 mM $CaCl_2$.

zGFRα1a$_{\Delta C}$-proteinA expressed alone or with zGDNF$_{\Delta N}$ in Hi5 cells was purified from the media using 0.5 ml IgG-sepharose resin per l culture and washed with ECD buffer with no added calcium (as described in section 7.4.2.1). Samples of washed IgG-sepharose resin bound zGFRα1a$_{\Delta C}$-proteinA and zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$-proteinA were run on SDS-PAGE to confirm similar levels of zGFRα1a$_{\Delta C}$-proteinA were present in both samples and that the zGDNF$_{\Delta N}$ and zGFRα1a$_{\Delta C}$-proteinA levels were approximately stoichiometric. The IgG-sepharose resin bound zGFRα1a$_{\Delta C}$-proteinA and zGDNF$_{\Delta N}$/zGFRα1a$_{\Delta C}$-proteinA were each divided into three aliquots containing 0.15 ml resin. 2× excess zRET-ECD ± 2 mM $CaCl_2$ or zRET-CLD(1–4) + 2 mM $CaCl_2$ were added to these aliquots and incubated at 4 °C for 1 hour with rotation. Unbound protein was then removed and the resin washed with 10 ml ECD buffer ± 2 mM $CaCl_2$. Bound protein was then eluted from the resin in ECD buffer ± 2 mM $CaCl_2$ by cleavage of the zGFRα1a$_{\Delta C}$ Protein A tag with GST tagged 3C-protease 100 μg per 1 ml resin. Eluted protein samples were run on SDS-PAGE to determine whether zRET-ECD or zRET-CLD(1–4) had bound.

### 7.6.9 Testing glycosidases assay

The protein glycosidase enzymes endoglycosidase F1 (Endo-F1; Calbiochem), endoglycosidase H (Endo-H; Sigma Aldrich) and peptide N-glycosidase F (PNGase F; produced by the Structural Biology Laboratory) were tested for their ability to deglycosylate zRET-ECD at 4 °C, 20 °C and 37 °C over 20 hours. The 50 μl assay mixture contained 24 μg zRET-ECD and >2 units of glycosidase in ECD buffer with calcium supplemented with the glycosidase buffer (1× final concentration; Table 7.7). A control reaction without glycosidase was also carried out at each temperature. 12 μl gel

samples were removed from each assay mixture after 1, 3 and 20 hours. Gel samples were run on non-reducing SDS-PAGE for analysis.

| 10 × Buffer | Components |
|---|---|
| Endo-F1 | 500 mM Sodium phosphate buffer pH 5.5 |
| Endo-H | 500 mM Sodium citrate buffer pH 5.5 |
| PNGase F | 500 mM Sodium phosphate buffer pH 7.5 |

**Table 7.7 Glycosidase enzyme buffers**

### 7.6.10 RET tyrosine kinase activity assay

pRET-KD wild type and mutant kinase activity was measured using a continuous ADP-coupled assay, where ATP consumption is coupled to NADH oxidation (Knowles *et al.*, 2006). The reaction mix is detailed in Table 7.8. A poly($E_4Y$) peptide of random length (Sigma Aldrich) was used as a substrate and its concentration was varied from 0.25–8 mg/ml, which allowed calculation of the kcat and the Km for the peptide since both wild type and mutant RET-KDs in these conditions reached saturation at < 8 mg/ml peptide substrate. Each substrate concentration was tested in triplicate for each of the RET-KD proteins. RET-KD and poly($E_4Y$) peptide were pre-incubated for 10 minutes. The reaction was then started by addition of ATP with the other assay components. 100 µl reactions were carried out in triplicate in a 96 well format. Loss of NADH was followed by measurement of the Abs343 every minute for 40 minutes in a Tecan plate reader. Addition of ADP to the reaction mix was used as a positive control to ensure assay components were functional. Peptide was left out of the mix to measure background ATP consumption, which was very low. The initial rate of reaction for each peptide concentration was calculated from the linear portion of the Abs343 curve, and the background ATP consumption reaction rate was deducted. Kinetic constants were calculated by fitting the data to the Michaelis-Menten equation using Prism® software (GraphPad).

The IC50 for sunitinib malate (Toronto Research Chemicals) was calculated for phosphorylated and non-phosphorylated RET-KD using this assay with 1 mg/ml poly($E_4Y$) peptide and varying the final sunitinib malate concentration in the assay mix from 0–100 µM. Each sunitinib concentration was tested in triplicate for both proteins. RET-KD was pre-incubated with sunitinib malate for 30 minutes prior to addition of

peptide and then other assay components. Otherwise reactions were carried out and data analysed as detailed above.

| Final concentration | Component |
|---|---|
| 25 µg/ml (740 nM) | RET-KD |
| 0.25–8 mg/ml | poly($E_4Y$) peptide |
| 100 mM | Tris pH 7.5 |
| 10 mM | $MgSO_4$ |
| 2 mM | Phosphoenolpyruvate |
| 1 mM | Reduced nicotinamide adenine dinucleotide (NADH) |
| 1 mM | ATP pH 8 |
| 2 units | Pyruvate kinase (Calbiochem) |
| 1.4 units | Lactate dehydrogenase (Calbiochem) |

**Table 7.8 Kinase activity assay reaction mix**

## 7.7 Homology models of RET and GFRα subdomains

To generate accurate homology models for individual zRET CLDs (for SAXS) and hGFRα1-D1 (for EM) the closest structural homologues from the PDB were identified and structure prediction based on these homologues was carried out using PHYRE2 (Kelley and Sternberg, 2009). From the input amino acid sequence PHYRE2 runs the secondary structure prediction programs Psi-Pred (McGuffin *et al.*, 2000), SSPro (Pollastri *et al.*, 2002) and JPred (Cole *et al.*, 2008), and generates a consensus secondary structure prediction, from which structural homologues of the input protein are identified and a tertiary structure prediction is made. The 'normal' mode of PHYRE2 was used for zRET-CLD3 and hGFRα-D1 structure prediction. In this mode PHYRE2 identifies structural homologues from the PDB of the input and produces a tertiary structure prediction model of the input for each identified PDB entry. This model is generated based primarily on secondary structure alignment, and regions of the input that do not align are left out of the model.

For hGFRα1-D1, the GFRα3-D3 from PDB 2GH0 was the best identified homologue (92.7 % confidence) and was therefore chosen as the template for the hGFRα1-D1 structure prediction model. zRET-CLD3 was more difficult to model. The best homologue was a cadherin domain from C-cadherin (PDB 1L3W) with 69.4% confidence. However, the structure prediction model using 1L3W as a template placed

some of the calcium-binding residues that contribute to the CLD2–CLD3 junction at the wrong end of the CLD3 model. After looking through the other PHYRE2 identified homologues, the model based on human protocadherin 9 (PDB 2EE0) had a better beta strand topology for the second half of CLD3. These two models were therefore combined using COOT (Emsley and Cowtan, 2004), basing the arrangement of calcium-binding residues on C-cadherin (PDB 1L3W; alignment of C-cadherin to RET CLDs is shown in Figure 1.7).

The 'intensive' mode was used for the zRET-CLD(1–2) and zRET-CLD4 models. This mode builds a model for the entire input amino acid sequence. As in the 'normal' mode structural homologues from PDB are identified; but in the 'intensive' mode if a number of structures with high probability that the input and identified structure are homologues, these structures used jointly as templates for tertiary structure modelling of the input. For zRET-CLD(1–2) the hRET-CLD(1–2) (PBD 2X2U) was the sole template used for structure prediction. In this manner 96% of zRET-CLD(1–2) could be modelled with >90% confidence. zRET-CLD4 was modelled based on four templates: two different cadherin domains from 1NCI (N-cadherin) and individual domains from 2X2U (hRET-CLD(1–2)) and 1OP4 (N-cadherin prodomain). The probability that these templates were homologues was between 87 and 89%. This generated a zRET-CLD4 model where ~90% was modelled with >80% confidence. Any regions of the input that could not be modelled by alignment were modelled *ab initio* using POING (Kelley and Sternberg, 2009).

The 'intensive' mode did not generate satisfactory models of zRET-CLD3 and hGFRα1-D1 since a larger proportion of these structures were modelled *ab initio* in this mode.

### 7.7.1 Glycosylating structure prediction models

The GLYCAM web server (Woods Group, 2005-2012) was used to add N-linked glycans to the structure prediction models (Kirschner *et al.*, 2008). Paucimannose moieties (A-D-Manp-(1-6)-[A-D-Manp-(1-3)]-B-D-Manp-(1-4)-B-D-GlcpNAc-(1-4)-[A-L-Fucp-(1-6)]-B-D-GlcpNAc) were added to the predicted N-linked glycosylation sites, followed by energy minimisation with AMBER (Case *et al.*, 2012).

## 7.8 Small angle X-ray scattering (SAXS)

### 7.8.1 Sample preparation and data collection at the ESRF

Size exclusion purified zRET-ECD and hRET-ECD were used in SAXS experiments at the European Synchrotron Radiation Facility (ESRF). SAXS data was collected on the bioSAXS beamline ID14-3 with a Pilatus 1M detector (Dectris Ltd.) and a sample-detector distance of 2.5 m. Samples were passed through a 1.8 mm diameter quartz capillary (mounted in vacuum) for data collection. Data was collected for a range of concentrations between 2–7 mg/ml for each protein to assess whether interparticle effects such as aggregation were present. 10 frames of 2 second exposures were collected for each sample whilst the sample was continually moving through the capillary. The SAXS data measured on a 2D detector for each frame was converted to a 1D SAXS profile of intensity of the scattered beam (I) as a function of the modulus of the scattering vector ($s = (4\pi/\lambda)\sin\theta$, where $\lambda$ is the wavelength (Å) and $2\theta$ the scattering angle). Comparison between the 1D SAXS profile of each frame allowed any radiation damage that occurred over the course of data collection to be detected. The frames showing the same SAXS profile were merged in order to improve the signal to noise ratio. Data was also collected for the buffers and this was subtracted from the protein sample data to remove the contribution from background scatter from the buffer and experimental setup.

### 7.8.2 Sample preparation and data collection at SOLEIL

SAXS data collection at Synchrotron SOLEIL at the SWING beamline was carried out using their online High Performance Liquid Chromatography (HPLC) setup (David and Pérez, 2009), which allows separation of any aggregates/oligomers present in the sample from the monomer peak directly before SAXS data collection. The beamline is equipped with a PCCD170170 detector (AVIEX) and a sample-detector distance of 1.8 m was used. Data was collected from samples passing through a thermostated 1.5 mm diameter quartz capillary within the vacuum chamber. Each frame of data collected corresponded to a 1.5 or 2 second exposure, which was followed by 1 second dead time.

Size-exclusion purified zRET and zTC samples at 12–24 mg/ml were injected onto a 5 ml SHODEX KW403-4F column size exclusion column connected to a HPLC system (Agilent) and eluted directly into the data collection capillary. Size exclusion was carried out in ECD buffer at a flow rate of 150 µl/min. Around 60 frames of SAXS data was collected for the size exclusion buffer, after the void volume of the column (5 ml) but before protein elution. SAXS data was collected for the protein sample across the entire size exclusion elution peak, often corresponding to 100 frames of SAXS data (as determined by the Abs280 trace).

#### *7.8.2.1 HPLC coupled SAXS initial data processing*

Raw image SAXS data collected at SOLEIL was processed using Foxtrot (SOLEIL in-house software) to generate the 1D SAXS profiles for each data collection frame. The frames of data from the buffer were then merged and subtracted from each individual frame of data from the HPLC protein elution peak. The most intense SAXS profile (corresponding to the data collected at the centre of the HPLC peak) was then analysed in Foxtrot to determine the Rg using the Guiner approximation ($I(s) = I(0)\exp(-s^2Rg^2/3)$ where s is the modulus of the scattering vector, I(0) is the intensity extrapolated to s = 0, and Rg is the radius of gyration) (Guinier and Fournet, 1955). Using the same limits for the Guiner approximation the Rg was then calculated for every data collection SAXS profile across the HPLC peak. This analysis allowed any changes in the Rg across the peak to be detected. Several frames (7–10) corresponding to the centre of the HPLC peak, which were determined to have the same Rg, were then merged and the resulting SAXS profile was used for subsequent analysis.

### 7.8.3 SAXS data processing

The final averaged experimental SAXS profiles from ERSF and SOLEIL data collections were processed using the ATSAS suite (Konarev *et al.*, 2006). Firstly PRIMUS was used to determine the Rg using the Guiner approximation (Guinier and Fournet, 1955) and generate Kratky plots (Kratky and Porod, 1949) to confirm the proteins were folded (Putnam *et al.*, 2007). GNOM (Svergun, 1992) was then used to fit the data and generate the pair distance distribution function (P(r)). GNOM uses a number of criteria to assess the quality of the output fit and P(r). These are the chi-

square value (or discrepancy between the experimental data ($I_{exp}$) and the fit ($I_{calc}$)), the smoothness of the solution, the stability of the solution, the presence of systematic deviations between the experimental data and the fit, the validity of the chosen real space interval and the positive part of the P(r) equalling 1 (Svergun, 1992). These five criteria are weighted and combined to give a score for the GNOM solution of between 0 and 1. The maximum vector length (Dmax) in the sample was determined iteratively in GNOM, using 3× Rg as the initial estimate. The Dmax was chosen based on the score of the GNOM solution and whether the P(r) equalled zero at 0 Å and the Dmax. The final scores for the GNOM solutions are detailed in Table 7.9.

| Protein | GNOM solution score | Quality of solution |
|---|---|---|
| **ERSF SAXS data** | | |
| hRET-ECD | 0.724 | Reasonable |
| zRET-ECD | 0.601 | Reasonable |
| | | |
| **Soleil HPLC-coupled SAXS data** | | |
| zRET-CLD(1–4) | 0.704 | Reasonable |
| Partially-deglycosylated zRET-CLD(1–4) | 0.799 | Good |
| zRET-ECD | 0.724 | Reasonable |
| Partially-deglycosylated zRET-ECD | 0.840 | Good |
| Calcium-free zRET-ECD | 0.663 | Reasonable |
| Calcium-free zRET-CLD(1–4) | 0.666 | Reasonable |
| zTC | 0.700 | Reasonable |

**Table 7.9 Quality of GNOM pair distance distribution function solutions**

The GNOM solution score is a composite score based on five criteria, as described in the text. The quality of the solution is defined by GNOM as reasonable where the score is > 0.5 and good when it is > 0.75.

The SAXS MoW web server (Fischer *et al.*, 2009) was used on the GNOM output file to estimate the molecular weight of the sample from the Kratky plot of the SAXS data. This analysis confirmed the oligomeric state of the protein in the X-ray beam. The molecular weight of the sample can also be calculated from the I(0) if the sample concentration is known. However the exact concentration of the sample in the beam when using the online HPLC is unknown.

### 7.8.4 *Ab-initio* shape determination from SAXS data

DAMMIF (Franke and Svergun, 2009) was used to generate *ab-initio* models of the shape of the protein from the GNOM output file. DAMMIF was run multiple times (10–50) and the resulting models were aligned and an averaged model generated using DAMAVER (Volkov and Svergun, 2003) in batch mode.

### 7.8.5 Building domain arrangement models based on SAXS data

Interpretation of the SAXS data was carried out primarily in CHIMERA (Goddard *et al.*, 2007). The structure prediction models of zRET-CLD(1–3) and zRET-CLD4 were docked into the average SAXS *ab initio* models using CHIMERA's 'fit-in-map' tool. Since SAXS data does not determine the handedness of the protein the mirror image of each *ab initio* model was generated in IMAGIC (van Heel *et al.*, 1996). The zRET-CLD models were also docked into the enantiomorph *ab initio* models and the final *ab initio* model chosen based on the fit of the zRET-CLD(1–3) model.

Since structure prediction models for the entire glycosylated zRET-CLD(1–4)—except the linker between CLD3 and CLD4—could be generated, the arrangement of the domains as determined by fitting into the *ab initio* envelope could then be compared to the experimental SAXS profile. The theoretical SAXS curve of the fitted glycosylated zRET-CLD(1–4) was generated and compared to the experimental data using the FoXS server (Schneidman-Duhovny *et al.*, 2010) implemented in CHIMERA. Different CLD(1–4) models were compared on the basis of the discrepancy between the theoretical SAXS curve of the model to the experimental data, as determined by the chi-square value ($\chi^2 = \Sigma((I_{exp}(s) - I_{model}(s))/\sigma_{exp}(s))$ calculated over s from 0–0.5 $A^{-1}$).

### 7.8.6 Rigid-body modelling against SAXS data

Preliminary rigid body modelling of CLD(1-4) against the glycosylated zRET-CLD(1-4) data was carried out by Dr. Andrew Purkiss-Trew using SASREF (Petoukhov and Svergun, 2005). This analysis was carried out using the hRET-CLD(1-2) structure and individual C-cadherin EC domains as CLD3 and CLD4 representatives. The best-fit arrangement of the input domains to the SAXS data, as measured by the chi-square value, was carried out by SASREF with the following imposed constraints on the

relationship between the three structural models. These constraints were designed to ensure the correct ends of the CLD models were contacting in the final model. Residues at the base of CLD(1-2) and the top of the CLD3 model were limited to being no more than 10 Å apart, and residues at the base of the CLD3 model and the top of CLD4 were limited to being no more than 7 Å apart. The final model from this simulated annealing procedure using these unglycosylated models had a chi-square value of 11.4 and showed a ~150° angle between CLD2 and CLD3 and a ~100° angle between CLD3 and CLD4, consistent with the arrangement of the CLDs from fitting in the deglycosylated zRET-ECD *ab initio* envelope.

## 7.9 X-ray crystallography

### 7.9.1 Co-crystallisation with inhibitor and peptide

The pRET-KD was co-crystallised with inhibitors (sunitinib and AMPPNP) and a peptide (RET activation loop peptide (ALpep): SRDVYEEDSYVKRSQG; produced by the Cancer Research UK London Research Institute Peptide Synthesis Laboratory). This was carried out by incubating the affinity or size exclusion purified pRET-KD with inhibitor at a 1:3 protein:inhibitor molar ratio on ice for 1 hour prior to concentration of the protein for crystallisation screening. The ALpep was added to pRET-KD:AMPPNP following concentration to 3–4 mg/ml. The ALpep was added at an approximate 1:3 pRET-KD:peptide molar ratio and allowed to bind for 1 hour on ice, prior to setting up the crystallisation screens.

### 7.9.2 Initial crystallisation screening

Initial crystallisation screening was carried out using 96 well format kits from Hampton Research, Molecular Dimensions, Qiagen, Emerald Biosciences and Jena Biosciences, using the vapour diffusion method. Screening was carried out in 2-drop 96 well MRC plates at either 4 °C or 20 °C. 100 µl of kit solutions were transferred to the MRC plate using a Hydra robot (Thermo Scientific). Sitting drops of typically 100 nl of the purified protein at 3–30 mg/ml plus 100 nl of the well solution were aliquoted using a Mosquito robot (TTP Labtech). Screens were then incubated in a temperature controlled incubator or a robotic crystal hotel and imager (Rock Imager, Formulatrix). The concentration of

the protein used in crystallisation screening was varied based on the relative levels of precipitated versus clear drops observed in the previously setup screens, with the protein concentration reduced if drops were predominantly precipitated and increased if they were predominantly clear.

If no crystals were observed in a range of crystal screens and the protein concentration had been optimised, the protein buffer was then varied and optimised. Low (50 mM) and high (400 mM) NaCl were typically tried. A thermal stability shift assay was also conducted for zRET-ECD (Ericsson *et al.*, 2006): Different buffers with varying pH (6–8) and NaCl concentration (25–500 mM) and ± 5% glycerol were tested for their effect on the thermal stability of zRET-ECD. Sypro Orange dye (Thermofluor) was added to 5 µg zRET-ECD in these buffers in transparent PCR tubes with a total reaction volume of 100 µl. The temperature was raised from 20–90 °C in a real-time PCR machine (iQ5, Biorad) and the Sypro Orange fluorescence followed. Protein denaturation was associated with an increased Sypro Orange fluorescence and therefore the melting temperature (Tm) could be determined from the inflection point of the Sypro Orange fluorescence curve. The buffer in which zRET-ECD had the highest Tm was 20 mM Tris pH7, 200 mM NaCl, 1 mM $CaCl_2$, 5% glycerol. The presence of glycerol was therefore also tried in crystallisation screening of zRET-ECD.

### 7.9.3 Crystal optimisation

If crystals were observed in the crystallisation screens and were confirmed to be protein either by UV absorption (Rock Imager) or X-ray diffraction, the crystallisation condition was reproduced and optimised manually using a larger sitting drop size (1–5 µl) in a 24 well format tray. Typically this involved varying the pH, salt concentration, precipitant concentration and protein concentration around that of the initial condition from the screening kit. The aim was to produce three-dimensional single crystals that were as large as possible. In the case of pRET-KD:sunitinib an additive screen was also carried out using the additive screen kit from Hampton Research.

### 7.9.4 Cryo-protection and crystal mounting

In preparation for X-ray data collection crystals were cryoprotected, mounted in litholoops (Molecular Dimensions) and flash-frozen in liquid nitrogen. The cryoprotectant solution typically contained the crystallisation condition with additional glycerol or PEG400. The cryoprotectant used was based on McFerrin and Snell, 2002 and tested by applying to a larger litholoop than intended for the crystal, freezing in the cryo-stream and taking a test diffraction image with 5 minute exposure time on the in-house X-ray system (Rigaku). If no ice rings were observed in the diffraction pattern, the cryoprotectant was used for crystal cryoprotection.

### 7.9.5 Diffraction data collection

X-ray diffraction data were collected at Diamond Light Source or ESRF. Details of the beamline used are given in the results section where the X-ray data is presented. X-ray diffraction data collection was carried out as follows. The loop was mounted on the goniometer and the crystal was centred in the X-ray beam. The data collection strategy was determined using MOSFLM (Powell, 1999) or EDNA (Diamond Light Source) based on test shots at commonly 0°, 45° and 90° rotation angles of the goniometer.

### 7.9.6 Diffraction data processing

Diffraction data were processed using MOSFLM (Battye *et al.*, 2011; Powell, 1999), XDS (Kabsch, 2010) or D*TREK (Pflugrath, 1999). D*TREK was used for data collected on a Pilatus detector, since it worked better than MOSFLM with low intensity spots. These programs were used following the standard workflow as defined in their graphical user interfaces. The diffraction spots on the collected images were found, indexed and integrated based on refined unit cell parameters and crystal orientation, and the Laue and space group predicted. The integrated spot intensity data was then scaled and merged either in SCALA (Evans, 2006), with the Laue and space group determined using POINTLESS (Evans, 2006), or directly in XDS or D*TREK (SCALA and POINTLESS are part of the CCP4 suite (Collaborative Computational Project, 1994)). The output files generated by these scale and merge programs detailing data quality metrics against resolution and image number were analysed to determine the resolution

limit, and to identify any poor images to be discarded. The data were then commonly scaled and merged again at the determined resolution with the poor images removed.

### 7.9.7 Molecular replacement

The pRET-KD:sunitinib and pRET-KD:adenosine structures were solved by molecular replacement using the published pRET-KD:PP1 structure (PDB 2IVV). The number of molecules in the asymmetric unit was determined using the 'Cell Content Analysis' tool, which determines the Matthew's Coefficient, in the CCP4 suite. For both pRET-KD:sunitinib and pRET-KD:adenosine there was found to be one pRET-KD molecule in the asymmetric unit. Molecular replacement was carried out using the maximum-likelihood program PHASER (McCoy *et al.*, 2007) with just the protein chain coordinates from the pRET-KD:PP1 structure. This process yielded one good solution for both pRET-KD:sunitinib and pRET-KD:adenosine.

#### *7.9.7.1 Determination of the pRET-KD$_{P766R/Q910R}$ arrangement in the crystal*

The pRET-KD$_{P766R/Q910R}$ structure could not be solved by molecular replacement, as the resolution of the dataset was too low (~7 Å). Since an unpublished structure of the RET-KD at 3.3 Å resolution had been solved in the lab with the same space group and very similar unit cell dimensions, we reasoned that the arrangement of the RET-KD molecules in the pRET-KD$_{P766R/Q910R}$ crystal might be the same as in this structure. To test this we used the protein coordinates of the 3.3 Å structure with cell dimensions altered to match those of pRET-KD$_{P766R/Q910R}$. These coordinates were subjected to 10 rounds of restrained refinement against the scaled and merged pRET-KD$_{P766R/Q910R}$ dataset in REFMAC (Murshudov *et al.*, 1997). This generated a very low-resolution but connected electron density map with the model having an $R_{free}$ of 43 %.

### 7.9.8 Refinement and model building

Alternate rounds of refinement and model building were carried out to the molecular replacement solution electron density map and protein coordinates. This process was conducted to improve the phases of the data and thereby improve the electron density map. The pRET-KD:sunitinib structure was refined in REFMAC (Murshudov *et al.*,

1997) and the pRET-KD:adenosine structure was refined in PHENIX (Adams *et al.*, 2010) and BUSTER-TNT (Global Phasing Ltd.) (Blanc *et al.*, 2004). In this case BUSTER-TNT was successful, and improved the connectivity of the electron density map for the unmodelled loop, allowing it to be built confidently. Model building for both structures was carried out in COOT (Emsley and Cowtan, 2004). TLS groups were used in refinement of the structures, allowing individual rigid body refinement of the RET-KD N and C-lobes. The refinement and model building cycle was continued until the $R_{free}$ ceased to improve and the electron density map was as well interpreted by the structural model as possible whilst maintaining a difference between the $R_{work}$ and $R_{free}$ of less than 7%. BUSTER-TNT was used in the latter stages of pRET-KD:adenosine refinement to improve the density for the split conformation of the gly-rich loop, since it was designed to show less bias towards the partial model in refining the phases, thereby improving the electron density for the unmodelled parts of the structure. Structure validation was carried out using the tools in COOT.

## 7.10 Negative stain electron microscopy

### 7.10.1 Preparation of the EM grids and data collection

Size exclusion purified mTC and zTC were used for negative stain EM. Sample purity and complex stoichiometry was assessed by SDS-PAGE and, for the mTC, a TapeStation (Agilent), which allows quantification of the relative levels of the complex components. In the case of the zTC crosslinking was required for complex integrity. Crosslinking was carried out by addition of 0.05% glutaldehyde to the sample and incubation on ice for 3 minutes prior to placing the sample on the EM grid.

The samples were applied to glow-discharged (PELCO EasiGlow) quantifoil (R1.2/1.3, Quantifoil Microtool GmBH) grids coated with a thin carbon layer and stained with 2% (w/v) uranyl acetate. The grids were screened on a Tecnai T12 Spirit (FEI) operating at 120kV, and the grids with satisfactory sample quality and staining were stored. The images for analysis were collected with a FEI Tecnai F20 electron microscope operating at 200kV on a Tietz 4000×4000 CCD detector at a nominal magnification of 80000× corresponding to 2.17 Å/pixel.

### 7.10.2 Particle picking

Particle images were picked manually from the collected EM grid images using Boxer, part of the EMAN suite (Ludtke *et al.*, 1999). A box size of 192 × 192 pixels was used equating to 416.64 $Å^2$, between 2–3× the size of the particles. Particles were only picked from images of well-stained areas of the grid, where particles were well defined and lighter than the background. Particles were selected on the basis of their size—very large or very small particles compared to the intact complex were ignored. In total ~11000 particles were picked. Initial attempts at a 3D reconstruction from this stack of particle images were unsuccessful. Since it was observed that there was a wide range of staining levels in the particle image stack, particle picking was repeated with more stringent criteria for particle selection, with only the very well defined particles being selected. ~2000 particles were picked in this process. This approach did bias the particle image stack towards certain views, which were generally better defined. However, coverage of the range of possible views, as determined by the coverage of the Euler sphere during the subsequent 3D reconstruction process, was maintained.

The particle images were coarsened to 4.34 Å/pixel to increase the speed of subsequent computational processing steps. These raw images were then band-pass filtered, normalised and the unwanted background removed by masking. The high-pass filtering (200 Å), to suppress disturbing low spatial frequencies that can interfere with particle alignment procedures (van Heel *et al.*, 2000), was carried out in IMAGIC (van Heel *et al.*, 1996). Normalisation and application of a circular mask with soft edges was also carried out in IMAGIC. Low-pass filtering (30 Å) in SPIDER (Frank *et al.*, 1996) was then carried out to suppress the noisy high frequency data.

### 7.10.3 Alignment and classification

In order to improve the signal to noise ratio particle images are aligned, classified and class members combined and averaged. Initial characterisation of the data was carried out by ‘reference-free’ particle alignment and classification in EMAN (Ludtke *et al.*, 1999). During the 3D reconstruction process particle alignment was carried out in SPIDER (Frank *et al.*, 1996) using forward projections of the latest 3D reconstruction as references. Early in the 3D reconstruction process, forward projections corresponding

to 12° steps were generated and those with clear features were used as references for alignment. Once the 3D reconstruction had improved, a larger number of forward projections from a finer sampling of the Euler sphere were used as references for alignment to improve the detail in the resulting class averages. Classification and generation of class averages from the SPIDER aligned particle image stack was carried out in IMAGIC (van Heel *et al.*, 1996) using a multi-variate statistical analysis (MSA) method (van Heel, 1984).

### 7.10.4 3D reconstruction

#### *7.10.4.1 Angular reconstitution*

3D reconstruction was carried out using the angular reconstitution method (van Heel, 1987) as implemented in IMAGIC (van Heel *et al.*, 1996). This is a 'common-lines' based method. The common lines projection theorem states that any two 2D projections of the same 3D object will share a common 1D line projection. Since the particle images (and class averages) are 2D projections of the 3D molecule, this theorem applies, and the relative Euler-angle orientations between class averages can therefore be determined by identifying the common lines (van Heel *et al.*, 2000).

MSA of the centred particle stack in IMAGIC involves the generation of eigenimages, from which any symmetry present in the particle can be identified (van Heel *et al.*, 2000). The mTC was determined to have two-fold symmetry in this manner, which was consistent with the apparent two-fold evident in some of the particle images. Therefore the 3D reconstruction process was carried out with C2 symmetry imposed. In the absence of symmetry, at least three 2D projections are required to solve the orientation problem using the common lines method. In the presence of two-fold symmetry only two approximately orthogonal 2D projections (class averages) are required to solve the orientation problem. The 3D reconstruction of the mTC was therefore initiated using two representative class averages: One possessing two-fold symmetry, and therefore perpendicular to the two-fold axis, and the other apparently orthogonal to the first. Upon assignment of the relative Euler angles of these two class averages, these classes were then used as references to assign the Euler angles of the full range of class averages.

Whether the Euler angle assignment of each class average was correct was determined by generating a 3D reconstruction from the Euler angle assigned class averages, and then generating re-projections of this 3D corresponding to each class average in IMAGIC. Euler angle assignment was determined to be correct if the class average and the corresponding 3D re-projection matched. The correctly assigned class averages were then used as references for reassigning the Euler angles of the full range of class averages. This approach was carried out iteratively until no additional class averages could be correctly assigned.

At this stage, a 3D reconstruction was generated from the correctly Euler angle assigned class averages, and forward projections of this 3D reconstruction were used to realign and classify the particle image stack (section 7.10.3). The new class averages were then assigned Euler angles using the correctly assigned class averages from the previous round as references, and the iterative process of expanding the number of correctly assigned class averages repeated.

This cycle of iterative Euler angle assignment followed by realignment and classification of the particle image stack to the resulting 3D reconstruction was repeated. The aim was to improve the 3D reconstruction in each round by increasing the coverage of the Euler sphere with correctly assigned class averages and improving the detail in the class averages, by alignment of the particles to a better set of references.

Once no further improvement in either Euler sphere coverage or class average detail could be observed using this angular reconstitution method with the small (~2000 particle images) stringently selected particle image stack, the 3D reconstruction was refined by the projection matching procedure using the ~11000 particle image stack.

#### *7.10.4.2 Projection matching*

Projection matching was carried out in IMAGIC (van Heel *et al.*, 1996). In this method, matching the class averages to forward projections of the current 3D reconstruction is used for Euler angle assignment of the class averages. Otherwise the iterative procedure of 3D reconstruction from the correctly assigned classes, followed by generation of new class averages by alignment of the particle image stack to the 3D reconstruction and subsequent classification, is the same as the angular reconstitution method described

above. The goal of improving the Euler sphere coverage and detail in the class averages to produce a better 3D reconstruction is also the same.

### 7.10.5 Interpretation of the 3D reconstruction

The graphical program CHIMERA (Goddard *et al.*, 2007) was used for visualisation and interpretation of the final 3D reconstruction (EM map). The crystal structures and structure prediction models of the components of the complex were docked into the EM map both manually and using CHIMERA's 'fit-in-map' tool.

# Reference List

Abrescia, C., Sjostrand, D., Kjaer, S., and Ibanez, C.F. (2005). Drosophila RET contains an active tyrosine kinase and elicits neurotrophic activities in mammalian cells. FEBS Lett *579*, 3789-3796.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Airaksinen, M.S., Holm, L., and Hatinen, T. (2006). Evolution of the GDNF family ligands and receptors. Brain Behav Evol *68*, 181-190.

Airaksinen, M.S., and Saarma, M. (2002). The GDNF family: Signalling, biological functions and therapeutic value. Nat Rev Neurosci *3*, 383-394.

Airaksinen, M.S., Titievsky, A., and Saarma, M. (1999). GDNF family neurotrophic factor signaling: four masters, one servant? Mol Cell Neurosci *13*, 313-325.

Alvarado, D., Klein, D., and Lemmon, M. (2009). ErbB2 resembles an autoinhibited invertebrate epidermal growth factor receptor. Nature *461*, 287-291.

Amiel, J., Sproat-Emison, E., Garcia-Barcelo, M., Lantieri, F., Burzynski, G., Borrego, S., Pelet, A., Arnold, S., Miao, X., Griseri, P.*, et al.* (2008). Hirschsprung disease, associated syndromes and genetics: a review. J Med Genet *45*, 1-14.

Amoresano, A., Incoronato, M., Monti, G., Pucci, P., de Franciscis, V., and Cerchia, L. (2005). Direct interactions among Ret, GDNF and GFRalpha1 molecules reveal new insights into the assembly of a functional three-protein complex. Cell Signal *17*, 717-727.

Anders, J., Kjar, S., and Ibanez, C.F. (2001). Molecular Modeling of the Extracellular Domain of the RET Receptor Tyrosine Kinase Reveals Multiple Cadherin-like Domains and a Calcium-binding Site. J Biol Chem *276*, 35808-35817.

Arighi, E., Borrello, M.G., and Sariola, H. (2005). RET tyrosine kinase signaling in development and cancer. Cytokine Growth Factor Rev *16*, 441-467.

Asai, N., Fukuda, T., Wu, Z., Enomoto, A., Pachnis, V., Takahashi, M., and Costantini, F. (2006). Targeted mutation of serine 697 in the Ret tyrosine kinase causes migration defect of enteric neural crest cells. Development *133*, 4507-4516.

Asai, N., Iwashita, T., Matsuyama, M., and Takahashi, M. (1995). Mechanism of activation of the ret proto-oncogene by multiple endocrine neoplasia 2A mutations. Mol Cell Biol *15*, 1613-1619.

Bajetta, E., Guadalupi, V., and Procopio, G. (2009). Activity of Sunitinib in Patients With Advanced Neuroendocrine Tumors. J Clin Oncol *27*, 319-320.

Ballerini, P., Struski, S., Cresson, C., Prade, N., Toujani, S., Deswarte, C., Dobbelstein, S., Petit, A., Lapillonne, H., Gautier, E.F.*, et al.* (2012). RET fusion genes are associated with chronic myelomonocytic leukemia and enhance monocytic differentiation. Leukemia.

Baloh, R.H., Tansey, M.G., Johnson, E.M., Jr., and Milbrandt, J. (2000). Functional mapping of receptor specificity domains of glial cell line-derived neurotrophic factor (GDNF) family ligands and production of GFRalpha1 RET-specific agonists. J Biol Chem *275*, 3412-3420.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr *67*, 271-281.

Bespalov, M.M., Sidorova, Y.A., Tumova, S., Ahonen-Bishopp, A., Magalhaes, A.C., Kulesskiy, E., Paveliev, M., Rivera, C., Rauvala, H., and Saarma, M. (2011). Heparan sulfate proteoglycan syndecan-3 is a novel receptor for GDNF, neurturin, and artemin. J Cell Biol *192*, 153-169.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr D Biol Crystallogr *60*, 2210-2221.

Bocharov, E.V., Mineev, K.S., Volynsky, P.E., Ermolyuk, Y.S., Tkach, E.N., Sobol, A.G., Chupin, V.V., Kirpichnikov, M.P., Efremov, R.G., and Arseniev, A.S. (2008). Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J Biol Chem *283*, 6950-6956.

Boggon, T.J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B.M., and Shapiro, L. (2002a). C-Cadherin Ectodomain Structure and Implications for Cell Adhesion Mechanisms. Science *296*, 1308-1313.

Boggon, T.J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B.M., and Shapiro, L. (2002b). C-cadherin ectodomain structure and implications for cell adhesion mechanisms. Science *296*, 1308-1313.

Bonanomi, D., Chivatakarn, O., Bai, G., Abdesselem, H., Lettieri, K., Marquardt, T., Pierchala, Brian A., and Pfaff, Samuel L. (2012). Ret Is a Multifunctional Coreceptor that Integrates Diffusible- and Contact-Axon Guidance Signals. Cell *148*, 568-582.

Boulay, A., Breuleux, M., Stephan, C., Fux, C., Brisken, C., Fiche, M., Wartmann, M., Stumm, M., Lane, H.A., and Hynes, N.E. (2008). The Ret Receptor Tyrosine Kinase Pathway Functionally Interacts with the ER{alpha} Pathway in Breast Cancer. Cancer Res *68*, 3743-3751.

Bouyain, S., Longo, P.A., Li, S., Ferguson, K.M., and Leahy, D.J. (2005). The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. Proc Natl Acad Sci U S A *102*, 15024-15029.

Bugalho, M.J., Domingues, R., and Borges, A. (2009). A Case of Advanced Medullary Thyroid Carcinoma Successfully Treated with Sunitinib. Oncologist *14*, 1083-1087.

Cabrera, J.R., Manes, S., Sanchez Pulido, L., Rojas, A., Valencia, A., Manes, S., Naranjo, J., and Mellstrom, B. (2006). Gas1 is related to the glial cell-derived neurotrophic factor family receptors alpha and regulates Ret signaling. J Biol Chem *281*, 14330-14339.

Carlomagno, F., Anaganti, S., Guida, T., Salvatore, G., Troncone, G., Wilhelm, S.M., and Santoro, M. (2006). BAY 43-9006 inhibition of oncogenic RET mutants. J Natl Cancer Inst *98*, 326-334.

Carlomagno, F., Vitagliano, D., Guida, T., Ciardiello, F., Tortora, G., Vecchio, G., Ryan, A.J., Fontanini, G., Fusco, A., and Santoro, M. (2002). ZD6474, an orally available inhibitor of KDR tyrosine kinase activity, efficiently blocks oncogenic RET kinases. Cancer Res *62*, 7284-7290.

Carr, L.L., Mankoff, D.A., Goulart, B.H., Eaton, K.D., Capell, P.T., Kell, E.M., Bauman, J.E., and Martins, R.G. (2010). Phase II study of daily sunitinib in FDG-PET-positive, iodine-refractory differentiated thyroid cancer and metastatic medullary carcinoma of the thyroid with functional imaging correlation. Clin Cancer Res *16*, 5260-5268.

Carter, M.T., Yome, J.L., Marcil, M.N., Martin, C.A., Vanhorne, J.B., and Mulligan, L.M. (2001). Conservation of RET proto-oncogene splicing variants and implications for RET isoform function. Cytogenet Cell Genet *95*, 169-176.

Case, D.A., Darden, T.A., III, T.E.C., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M.*, et al.* (2012). AMBER 12 (University of California, San Francisco).

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I.*, et al.* (2007). Glycoprotein structural genomics: solving the glycosylation problem. Structure *15*, 267-273.

Chen, H., Ma, J., Li, W., Eliseenkova, A.V., Xu, C., Neubert, T.A., Miller, W.T., and Mohammadi, M. (2007). A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. Mol Cell *27*, 717-730.

Chen, H., Xu, C.F., Ma, J., Eliseenkova, A.V., Li, W., Pollock, P.M., Pitteloud, N., Miller, W.T., Neubert, T.A., and Mohammadi, M. (2008). A crystallographic snapshot of tyrosine trans-phosphorylation in action. Proc Natl Acad Sci U S A *105*, 19660-19665.

Cho, H.S., and Leahy, D.J. (2002). Structure of the extracellular region of HER3 reveals an interdomain tether. Science *297*, 1330-1333.

Chow, L.Q.M., and Eckhardt, S.G. (2007). Sunitinib: From Rational Design to Clinical Efficacy. J Clin Oncol *25*, 884-896.

Ciatto, C., Bahna, F., Zampieri, N., VanSteenhouse, H.C., Katsamba, P.S., Ahlsen, G., Harrison, O.J., Brasch, J., Jin, X., Posy, S.*, et al.* (2010). T-cadherin structures reveal a novel adhesive binding mechanism. Nat Struct Mol Biol *17*, 339-347.

Clayton, A.H., Walker, F., Orchard, S.G., Henderson, C., Fuchs, D., Rothacker, J., Nice, E.C., and Burgess, A.W. (2005). Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis. J Biol Chem *280*, 30392-30399.

Cohen, P. (2002). Protein kinases—the major drug targets of the twenty-first century? Nat Rev Drug Discov *1*, 309-315.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res *36*, W197-201.

Collaborative Computational Project, N. (1994). The CCP4 suite. Programs for protein crystallography. Acta crystallographica Section D *50*, 760-763.

Cordella, D., Muzza, M., Alberti, L., Colombo, P., Travaglini, P., Beck-Peccoz, P., Fugazzola, L., and Persani, L. (2006). An in-frame complex germline mutation in the juxtamembrane intracellular domain causing RET activation in familial medullary thyroid carcinoma. Endocr Relat Cancer *13*, 945-953.

Costantini, F., and Shakya, R. (2006). GDNF/Ret signaling and the development of the kidney. BioEssays *28*, 117-127.

Cowan-Jacob, S.W. (2006). Structural biology of protein tyrosine kinases. Cell Mol Life Sci *63*, 2608-2625.

Cowan-Jacob, S.W., Mobitz, H., and Fabbro, D. (2009). Structural biology contributions to tyrosine kinase drug discovery. Curr Opin Cell Biol *21*, 280-287.

D'Aniello, S., Irimia, M., Maeso, I., Pascual-Anaya, J., Jimenez-Delgado, S., Bertrand, S., and Garcia-Fernandez, J. (2008). Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. Mol Biol Evol *25*, 1841-1854.

Dar, A.C., Das, T.K., Shokat, K.M., and Cagan, R.L. (2012). Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. Nature *486*, 80-84.

Das, T., and Cagan, R. (2010). Drosophila as a novel therapeutic discovery tool for thyroid cancer. Thyroid *20*, 689-695.

David, G., and Pérez, J. (2009). Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. Journal of Applied Crystallography *42*, 892-900.

de Graaff, E., Srinivas, S., Kilkenny, C., D'Agati, V., Mankoo, B.S., Costantini, F., and Pachnis, V. (2001). Differential activities of the RET tyrosine kinase receptor isoforms during mammalian embryogenesis. Genes Dev *15*, 2433-2444.

Depetris, R.S., Hu, J., Gimpelevich, I., Holt, L.J., Daly, R.J., and Hubbard, S.R. (2005). Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. Mol Cell *20*, 325-333.

Dixit, A., Torkamani, A., Schork, N.J., and Verkhivker, G. (2009). Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: the impact on protein structure, dynamics, and stability. Biophys J *96*, 858-874.

Dojima, T., Nishina, T., Kato, T., Uno, T., Yagi, H., Kato, K., and Park, E.Y. (2009). Comparison of the N-linked glycosylation of human beta1,3-N-acetylglucosaminyltransferase 2 expressed in insect cells and silkworm larvae. J Biotechnol *143*, 27-33.

Donis-Keller, H., Dou, S., Chi, D., Carlson, K.M., Toshima, K., Lairmore, T.C., Howe, J.R., Moley, J.F., Goodfellow, P., and Wells, S.A., Jr. (1993). Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. Hum Mol Genet *2*, 851-856.

Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S.*, et al.* (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med *344*, 1031-1037.

Durbec, P., Marcos-Gutierrez, C.V., Kilkenny, C., Grigoriou, M., Wartiowaara, K., Suvanto, P., Smith, D., Ponder, B., Costantini, F., Saarma, M.*, et al.* (1996). GDNF signalling through the Ret receptor tyrosine kinase. Nature *381*, 789-793.

Eigenbrot, C., and Gerber, N. (1997). X-ray structure of glial cell-derived neurotrophic factor at 1.9 A resolution and implications for receptor binding. Nat Struct Mol Biol *4*, 435-438.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Endres, N.F., Engel, K., Das, R., Kovacs, E., and Kuriyan, J. (2011). Regulation of the catalytic activity of the EGF receptor. Curr Opin Struct Biol *21*, 777-784.

Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem *357*, 289-298.

Esseghir, S., Todd, S.K., Hunt, T., Poulsom, R., Plaza-Menacho, I., Reis-Filho, J.S., and Isacke, C.M. (2007). A role for glial cell-derived neurotrophic factor-induced expression by inflammatory cytokines and RET/GFRα1 receptor up-regulation in breast cancer. Cancer Res *67*, 11732-11741.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.

Ferguson, K.M., Berger, M.B., Mendrola, J.M., Cho, H.-S., Leahy, D.J., and Lemmon, M.A. (2003). EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. Mol Cell *11*, 507-517.

Fischer, H., de Oliveira Neto, M., Napolitano, H.B., Polikarpov, I., and Craievich, A.F. (2009). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. Journal of Applied Crystallography *43*, 101-109.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol *116*, 190-199.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342-346.

Furdui, C., Lew, E., Schlessinger, J., and Anderson, K. (2006). Autophosphorylation of FGFR1 kinase is mediated by a sequential and precisely ordered reaction. Mol Cell *21*, 711-717.

Gajiwala, K.S., Wu, J.C., Christensen, J., Deshmukh, G.D., Diehl, W., DiNitto, J.P., English, J.M., Greig, M.J., He, Y.A., Jacques, S.L.*, et al.* (2009). KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. Proc Natl Acad Sci U S A *106*, 1542-1547.

Garrett, T.P., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Kofler, M., Jorissen, R.N., Nice, E.C., Burgess, A.W.*, et al.* (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. Mol Cell *11*, 495-505.

Gavel, Y., and Heijne, G.v. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Engineering, Design and Selection *3*, 433-442.

Geneste, O., Bidaud, C., De Vita, G., Hofstra, R.M., Tartare-Deckert, S., Buys, C.H., Lenoir, G.M., Santoro, M., and Billaud, M. (1999). Two distinct mutations of the RET receptor causing Hirschsprung's disease impair the binding of signalling effectors to a multifunctional docking site. Hum Mol Genet *8*, 1989-1999.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J Struct Biol *157*, 281-287.

Grieco, M., Santoro, M., Berlingieri, M.T., Melillo, R.M., Donghi, R., Bongarzone, I., Pierotti, M.A., Della Ports, G., Fusco, A., and Vecchiot, G. (1990). PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. Cell *60*, 557-563.

Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B., and Chene, C. (1996). Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. Protein Sci *5*, 2617-2622.

Gschwind, A., Fischer, O.M., and Ullrich, A. (2004). The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat Rev Cancer *4*, 361-370.

Guinier, A., and Fournet, F. (1955). Small Angle Scattering of X-rays (New York: Wiley Interscience).

Gujral, T.S., Singh, V.K., Jia, Z., and Mulligan, L.M. (2006). Molecular Mechanisms of RET Receptor-Mediated Oncogenesis in Multiple Endocrine Neoplasia 2B. Cancer Res *66*, 10741-10749.

Gupta, R., Jung, E., and Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. http://www.cbs.dtu.dk/services/NetNGlyc/.

Hahn, M., and Bishop, J.M. (2001). Expression pattern of Drosophila ret suggests a common ancestral origin between the metamorphosis precursors in insect endoderm and the vertebrate enteric neurons. Proc Natl Acad Sci U S A *98*, 1053-1058.

Hancock, J.F. (2006). Lipid rafts: contentious only from simplistic standpoints. Nat Rev Mol Cell Biol *7*, 456-462.

Harrison, O., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.*, et al.* (2011). The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. Structure *19*, 244-256.

Harrison, O.J., Bahna, F., Katsamba, P.S., Jin, X., Brasch, J., Vendome, J., Ahlsen, G., Carroll, K.J., Price, S.R., Honig, B.*, et al.* (2010). Two-step adhesive binding by classical cadherins. Nat Struct Mol Biol *17*, 348-357.

Hatinen, T., Holm, L., and Airaksinen, M.S. (2007). Loss of neurturin in frog--Comparative genomics study of GDNF family ligand-receptor pairs. Mol Cell Neurosci *34*, 155-167.

Haussinger, D., Ahrens, T., Sass, H.J., Pertz, O., Engel, J., and Grzesiek, S. (2002). Calcium-dependent homoassociation of E-cadherin by NMR spectroscopy: changes in mobility, conformation and mapping of contact regions. J Mol Biol *324*, 823-839.

Heanue, T.A., and Pachnis, V. (2007). Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. Nat Rev Neurosci *8*, 466-479.

Heanue, T.A., and Pachnis, V. (2008). Ret isoform function and marker gene expression in the enteric nervous system is conserved across diverse vertebrate species. Mechanisms of Development *125*, 687-699.

Heldin, C.H. (1995). Dimerization of cell surface receptors in signal transduction. Cell *80*, 213-223.

Higgins, M.J., Forastiere, A., and Marur, S. (2009). New directions in the systemic treatment of metastatic thyroid cancer. Oncology (Williston Park) *23*, 768-775.

Hoffer, B., and Harvey, B. (2011). Is GDNF beneficial in Parkinson disease? Nature Reviews Neurology *7*, 600-602.

Hofstra, R.M., Landsvater, R.M., Ceccherini, I., Stulp, R.P., Stelwagen, T., Luo, Y., Pasini, B., Hoppener, J.W., van Amstel, H.K., Romeo, G.*, et al.* (1994). A mutation in the RET proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. Nature *367*, 375-376.

Hsu, T.A., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., Masuda, K., Whiteley, E.M., Fan, J.Q., Lee, Y.C., and Betenbaugh, M.J. (1997). Differential N-glycan patterns of secreted and intracellular IgG produced in Trichoplusia ni cells. J Biol Chem *272*, 9062-9070.

Hubbard, S.R. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J *16*, 5572-5581.

Hubbard, S.R., and Miller, W.T. (2007). Receptor tyrosine kinases: mechanisms of activation and signaling. Curr Opin Cell Biol *19*, 117-123.

Hubbard, S.R., Wei, L., Ellis, L., and Hendrickson, W.A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature *372*, 746-754.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell *109*, 275-282.

Ibanez, C.F. (2010). Beyond the cell surface: new mechanisms of receptor function. Biochem Biophys Res Commun *396*, 24-27.

Ichihara, M., Murakumo, Y., and Takahashi, M. (2004). RET and neuroendocrine tumors. Cancer Lett *204*, 197-211.

Jain, S., Naughton, C.K., Yang, M., Strickland, A., Vij, K., Encinas, M., Golden, J., Gupta, A., Heuckeroth, R., Johnson, E.M.*, et al.* (2004). Mice expressing a dominant-negative Ret mutation phenocopy human Hirschsprung disease and delineate a direct role of Ret in spermatogenesis. Development *131*, 5503-5513.

Jarvis, D.L., Richard, R.B., and Murray, P.D. (2009). Chapter 14 Baculovirus-Insect Cell Expression Systems. In Methods in Enzymology (Academic Press), pp. 191-222.

Jeanpierre, C., Mace, G., Parisot, M., Moriniere, V., Pawtowsky, A., Benabou, M., Martinovic, J., Amiel, J., Attie-Bitach, T., Delezoide, A.L.*, et al.* (2011). RET and GDNF mutations are rare in fetuses with renal agenesis or other severe kidney development defects. J Med Genet *48*, 497-504.

Jin, X., Walker, M.A., Felsovalyi, K., Vendome, J., Bahna, F., Mannepalli, S., Cosmanescu, F., Ahlsen, G., Honig, B., and Shapiro, L. (2012). Crystal structures of Drosophila N-cadherin ectodomain regions reveal a widely used class of Ca(2)+-free interdomain linkers. Proc Natl Acad Sci U S A *109*, E127-134.

Johnson, L.N. (2009). Protein kinase inhibitors: contributions from structure to clinical compounds. Q Rev Biophys *42*, 1-40.

Jura, N., Endres, N.F., Engel, K., Deindl, S., Das, R., Lamers, M.H., Wemmer, D.E., Zhang, X., and Kuriyan, J. (2009). Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. Cell *137*, 1293-1307.

Jura, N., Zhang, X., Endres, N.F., Seeliger, M.A., Schindler, T., and Kuriyan, J. (2011). Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. Mol Cell *42*, 9-22.

Kabsch, W. (2010). XDS. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E.*, et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Meth *5*, 53-55.

Kastner, J., Loeffler, H.H., Roberts, S.K., Martin-Fernandez, M.L., and Winn, M.D. (2009). Ectodomain orientation, conformational plasticity and oligomerization of ErbB1 receptors investigated by molecular dynamics. J Struct Biol *167*, 117-128.

Kato, M., Takeda, K., Kawamoto, Y., Iwashita, T., Akhand, A.A., Senga, T., Yamamoto, M., Sobue, G., Hamaguchi, M., Takahashi, M.*, et al.* (2002). Repair by Src kinase of function-impaired RET with multiple endocrine neoplasia type 2A mutation with substitutions of tyrosines in the COOH-terminal kinase domain for phenylalanine. Cancer Res *62*, 2414-2422.

Kawamoto, Y., Takeda, K., Okuno, Y., Yamakawa, Y., Ito, Y., Taguchi, R., Kato, M., Suzuki, H., Takahashi, M., and Nakashima, I. (2004). Identification of RET autophosphorylation sites by mass spectrometry. J Biol Chem *279*, 14213-14224.

Kelleher, F.C., and McDermott, R. (2008). Response to sunitinib in medullary thyroid cancer. Ann Intern Med *148*, 567.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc *4*, 363-371.

Khavrutskii, I., Grant, B., Taylor, S., and McCammon, J.A. (2009). A transition path ensemble study reveals a linchpin role for Mg(2+) during rate-limiting ADP release from protein kinase A. Biochemistry *48*, 11532-11545.

Kim, N., Stiegler, A., Cameron, T., Hallock, P., Gomez, A., Huang, J., Hubbard, S., Dustin, M., and Burden, S. (2008). Lrp4 is a receptor for Agrin and forms a complex with MuSK. Cell *135*, 334-342.

Kim, S., Tai, C.-Y., Mok, L.-P., Mosser, E., and Schuman, E. (2011). Calcium-dependent dynamics of cadherin interactions at cell-cell junctions. Proc Natl Acad Sci U S A *108*, 9857-9862.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., Gonzalez-Outeirino, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem *29*, 622-655.

Kjaer, S., Hanrahan, S., Totty, N., and McDonald, N.Q. (2010). Mammal-restricted elements predispose human RET to folding impairment by HSCR mutations. Nat Struct Mol Biol *17*, 726-731.

Kjaer, S., and Ibanez, C.F. (2003a). Identification of a surface for binding to the GDNF-GFR alpha 1 complex in the first cadherin-like domain of RET. J Biol Chem *278*, 47898-47904.

Kjaer, S., and Ibanez, C.F. (2003b). Intrinsic susceptibility to misfolding of a hot-spot for Hirschsprung disease mutations in the ectodomain of RET. Hum Mol Genet *12*, 2133-2144.

Kjaer, S., Kurokawa, K., Perrinjaquet, M., Abrescia, C., and Ibanez, C.F. (2006). Self-association of the transmembrane domain of RET underlies oncogenic activation by MEN2A mutations. Oncogene *25*, 7086-7095.

Klein, R.D., Sherman, D., Ho, W.H., Stone, D., Bennett, G.L., Moffat, B., Vandlen, R., Simmons, L., Gu, Q., Hongo, J.A.*, et al.* (1997). A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. Nature *387*, 717-721.

Knowles, P.P., Murray-Rust, J., Kjaer, S., Scott, R.P., Hanrahan, S., Santoro, M., Ibanez, C.F., and McDonald, N.Q. (2006). Structure and chemical inhibition of the RET tyrosine kinase domain. J Biol Chem *281*, 33577-33587.

Kodama, Y., Asai, N., Kawai, K., Jijiwa, M., Murakumo, Y., Ichihara, M., and Takahashi, M. (2005). The RET proto-oncogene: a molecular therapeutic target in thyroid cancer. Cancer Sci *96*, 143-148.

Kohno, T., Ichikawa, H., Totoki, Y., Yasuda, K., Hiramoto, M., Nammo, T., Sakamoto, H., Tsuta, K., Furuta, K., Shimada, Y.*, et al.* (2012). KIF5B-RET fusions in lung adenocarcinoma. Nat Med *18*, 375-377.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS2.1, a program package for small-angle scattering data analysis. Journal of Applied Crystallography *39*, 277-286.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography *36*, 1277-1282.

Kopec, J., and Schneider, G. (2011). Comparison of fluorescence and light scattering based methods to assess formation and stability of protein-protein complexes. J Struct Biol *175*, 216-223.

Kornev, A.P., Taylor, S.S., and Ten Eyck, L.F. (2008). A helix scaffold for the assembly of active protein kinases. Proc Natl Acad Sci U S A *105*, 14377-14382.

Kratky, O., and Porod, G. (1949). Roetgenuntersuchung Geloester Fadenmolekuele. . Recueil des Travaux Chimiques des Pays-Bas *68*, 1106-1122.

Kuriyan, J., and Eisenberg, D. (2007). The origin of protein interactions and allostery in colocalization. Nature *450*, 983-990.

Kutach, A.K., Villasenor, A.G., Lam, D., Belunis, C., Janson, C., Lok, S., Hong, L.N., Liu, C.M., Deval, J., Novak, T.J.*, et al.* (2010). Crystal structures of IL-2-inducible T cell kinase complexed with inhibitors: insights into rational drug design and activity regulation. Chem Biol Drug Des *76*, 154-163.

Lam, A.K., Montone, K.T., Nolan, K.A., and Livolsi, V.A. (1998). Ret oncogene activation in papillary thyroid carcinoma: prevalence and implication on the histological parameters. Hum Pathol *29*, 565-568.

Lanzi, C., Cassinelli, G., Nicolini, V., and Zunino, F. (2009). Targeting RET for thyroid cancer therapy. Biochem Pharmacol *77*, 297-309.

Lassila, J.K., Zalatan, J.G., and Herschlag, D. (2011). Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. Annu Rev Biochem *80*, 669-702.

Lemmon, M.A. (2009). Ligand-induced ErbB receptor dimerization. Exp Cell Res *315*, 638-648.

Lemmon, M.A., and Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. Cell *141*, 1117-1134.

Leppanen, V.M., Bespalov, M.M., Runeberg-Roos, P., Puurand, U., Merits, A., Saarma, M., and Goldman, A. (2004). The structure of GFRalpha1 domain 3 reveals new insights into GDNF binding and RET activation. EMBO J *23*, 1452-1462.

Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. Science *327*, 46-50.

Lipson, D., Capelletti, M., Yelensky, R., Otto, G., Parker, A., Jarosz, M., Curran, J.A., Balasubramanian, S., Bloom, T., Brennan, K.W., *et al.* (2012). Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med *18*, 382-384.

Liu, P., Cleveland, T., Bouyain, S., Byrne, P., Longo, P., and Leahy, D. (2012). A single ligand is sufficient to activate EGFR dimers. Proc Natl Acad Sci U S A *109*, 10861-10866.

Liu, X., Vega, Q.C., Decker, R.A., Pandey, A., Worby, C.A., and Dixon, J.E. (1996). Oncogenic RET receptors display different autophosphorylation sites and substrate binding specificities. J Biol Chem *271*, 5309-5312.

Liu, Y., and Gray, N.S. (2006). Rational design of inhibitors that bind to inactive kinase conformations. Nat Chem Biol *2*, 358-364.

Lopez-Ramirez, M.A., Dominguez-Monzon, G., Vergara, P., and Segovia, J. (2008). Gas1 reduces Ret tyrosine 1062 phosphorylation and alters GDNF-mediated intracellular signaling. Int J Dev Neurosci *26*, 497-503.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82-97.

Luo, Y., Tsuchiya, K.D., Il Park, D., Fausel, R., Kanngurn, S., Welcsh, P., Dzieciatkowski, S., Wang, J., and Grady, W.M. (2012). RET is a potential tumor suppressor gene in colorectal cancer. Oncogene.

Manning, G., Plowman, G.D., Hunter, T., and Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci *27*, 514-520.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002b). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Marcos-Gutierrez, C.V., Wilson, S.W., Holder, N., and Pachnis, V. (1997). The zebrafish homologue of the ret receptor and its pattern of expression during embryogenesis. Oncogene *14*, 879-889.

Martin, S.R., and Schilstra, M.J. (2008). Circular Dichroism and Its Application to the Study of Biomolecules. In Methods in Cell Biology, J.C. Dr. John, and Dr. H. William Detrich, III, eds. (Academic Press), pp. 263-293.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McFerrin, M.B., and Snell, E.H. (2002). The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. Journal of Applied Crystallography *35*, 538-545.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

McKern, N.M., Lawrence, M.C., Streltsov, V.A., Lou, M.-Z., Adams, T.E., Lovrecz, G.O., Elleman, T.C., Richards, K.M., Bentley, J.D., Pilling, P.A*., et al.* (2006). Structure of the insulin receptor ectodomain reveals a folded-over conformation. Nature *443*, 218-221.

Mendel, D.B., Laird, A.D., Xin, X., Louie, S.G., Christensen, J.G., Li, G., Schreck, R.E., Abrams, T.J., Ngai, T.J., Lee, L.B*., et al.* (2003). In vivo antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. Clin Cancer Res *9*, 327-337.

Mi, L.Z., Lu, C., Li, Z., Nishida, N., Walz, T., and Springer, T.A. (2011). Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. Nat Struct Mol Biol *18*, 984-989.

Mologni, L., Rostagno, R., Brussolo, S., Knowles, P.P., Kjaer, S., Murray-Rust, J., Rosso, E., Zambon, A., Scapozza, L., McDonald, N.Q*., et al.* (2010). Synthesis, structure-activity relationship and crystallographic studies of 3-substituted indolin-2-one RET inhibitors. Bioorg Med Chem *18*, 1482-1496.

Morandi, A., Plaza-Menacho, I., and Isacke, C.M. (2011). RET in breast cancer: functional and therapeutic implications. Trends Mol Med *17*, 149-157.

Mulligan, L.M., Kwok, J.B., Healey, C.S., Elsdon, M.J., Eng, C., Gardner, E., Love, D.R., Mole, S.E., Moore, J.K., and Papi, L. (1993). Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. Nature *363*, 458-460.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

Nagar, B., Bornmann, W.G., Pellicena, P., Schindler, T., Veach, D.R., Miller, W.T., Clarkson, B., and Kuriyan, J. (2002). Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). Cancer Res *62*, 4236-4243.

Niu, X.L., Peters, K.G., and Kontos, C.D. (2002). Deletion of the carboxyl terminus of Tie2 enhances kinase activity, signaling, and function. Evidence for an autoinhibitory mechanism. J Biol Chem *277*, 31768-31773.

Noble, M.E.M., Endicott, J.A., and Johnson, L.N. (2004). Protein Kinase Inhibitors: Insights into Drug Design from Structure. Science *303*, 1800-1805.

Nolen, B., Taylor, S., and Ghosh, G. (2004). Regulation of protein kinases: Controlling activity through activation segment conformation. Mol Cell *15*, 661-675.

Nollet, F., Kools, P., and van Roy, F. (2000). Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. J Mol Biol *299*, 551-572.

Nozaki, C., Asai, N., Murakami, H., Iwashita, T., Iwata, Y., Horibe, K., Klein, R.D., Rosenthal, A., and Takahashi, M. (1998). Calcium-dependent Ret activation by GDNF and neurturin. Oncogene *16*, 293-299.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M.*, et al.* (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell *110*, 775-787.

Oliver, A.W., Knapp, S., and Pearl, L.H. (2007). Activation segment exchange: a common mechanism of kinase autophosphorylation? Trends Biochem Sci *32*, 351-356.

Paratcha, G., Ledda, F., Baars, L., Coulpier, M., Besset, V., Anders, J., Scott, R., and Ibanez, C.F. (2001). Released GFRalpha1 potentiates downstream signaling, neuronal survival, and differentiation via a novel mechanism of recruitment of c-Ret to lipid rafts. Neuron *29*, 171-184.

Paratcha, G., Ledda, F., and Ibáñez, C.F. (2003). The neural cell adhesion molecule NCAM is an alternative signaling receptor for GDNF family ligands. Cell *113*, 867-879.

Parkash, V., and Goldman, A. (2009). Comparison of GFL-GFRalpha complexes: further evidence relating GFL bend angle to RET signalling. Acta Crystallogr Sect F Struct Biol Cryst Commun *65*, 551-558.

Parkash, V., Leppanen, V.M., Virtanen, H., Jurvansuu, J.M., Bespalov, M.M., Sidorova, Y.A., Runeberg-Roos, P., Saarma, M., and Goldman, A. (2008). The structure of the glial cell line-derived neurotrophic factor-coreceptor complex: insights into RET signaling and heparin binding. J Biol Chem *283*, 35164-35172.

Pelet, A., Geneste, O., Edery, P., Pasini, A., Chappuis, S., Atti, T., Munnich, A., Lenoir, G., Lyonnet, S., and Billaud, M. (1998). Various mechanisms cause RET-mediated signaling defects in Hirschsprung's disease. Journal Clin Invest *101*, 1415-1423.

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J *89*, 1237-1250.

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr *55*, 1718-1725.

Pines, G., Huang, P.H., Zwang, Y., White, F.M., and Yarden, Y. (2010). EGFRvIV: a previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism. Oncogene *29*, 5850-5860.

Plaza-Menacho, I., Burzynski, G.M., de Groot, J.W., Eggen, B.J., and Hofstra, R.M. (2006). Current concepts in RET-related genetics, signaling and therapeutics. Trends Genet *22*, 627-636.

Plaza-Menacho, I., Mologni, L., Sala, E., Gambacorti-Passerini, C., Magee, A.I., Links, T.P., Hofstra, R.M., Barford, D., and Isacke, C.M. (2007). Sorafenib functions to potently suppress RET tyrosine kinase activity by direct enzymatic inhibition and promoting RET lysosomal degradation independent of proteasomal targeting. J Biol Chem *282*, 29230-29240.

Plaza-Menacho, I., Morandi, A., Robertson, D., Pancholi, S., Drury, S., Dowsett, M., Martin, L.A., and Isacke, C.M. (2010). Targeting the receptor tyrosine kinase RET sensitizes breast cancer cells to tamoxifen treatment and reveals a role for RET in endocrine resistance. Oncogene *29*, 4648-4657.

Pokutta, S., Herrenknecht, K., Kemler, R., and Engel, J. (1994). Conformational changes of the recombinant extracellular domain of E-cadherin upon calcium binding. Eur J Biochem *223*, 1019-1026.

Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins *47*, 228-235.

Powell, H.R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallogr D Biol Crystallogr *55*, 1690-1695.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys *40*, 191-285.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol *333*, 721-745.

Runeberg-Roos, P., and Saarma, M. (2007). Neurotrophic factor receptor RET: structure, cell biology, and inherited diseases. Ann Med *39*, 572-580.

Runeberg-Roos, P., Virtanen, H., and Saarma, M. (2007). RET(MEN 2B) is active in the endoplasmic reticulum before reaching the cell surface. Oncogene *26*, 7909-7915.

Salvatore, D., Barone, M.V., Salvatore, G., Melillo, R.M., Chiappetta, G., Mineo, A., Fenzi, G., Vecchio, G., Fusco, A., and Santoro, M. (2000). Tyrosines 1015 and 1062 are in vivo autophosphorylation sites in ret and ret-derived oncoproteins. J Clin Endocrinol Metab *85*, 3898-3907.

Salvatore, D., Melillo, R.M., Monaco, C., Visconti, R., Fenzi, G., Vecchio, G., Fusco, A., and Santoro, M. (2001). Increased in Vivo Phosphorylation of Ret Tyrosine 1062 Is a Potential Pathogenetic Mechanism of Multiple Endocrine Neoplasia Type 2B. Cancer Res *61*, 1426-1431.

Santoro, M., Carlomagno, F., Romano, A., Bottaro, D.P., Dathan, N.A., Grieco, M., Fusco, A., Vecchio, G., Matoskova, B., and Kraus, M.H. (1995). Activation of RET as a dominant transforming gene by germline mutations of MEN2A and MEN2B. Science *267*, 381-383.

Santoro, M., Melillo, R.M., Carlomagno, F., Vecchio, G., and Fusco, A. (2004). Minireview: RET: normal and abnormal functions. Endocrinology *145*, 5448-5451.

Santoro, M., Melillo, R.M., and Fusco, A. (2006). RET/PTC activation in papillary thyroid carcinoma: European Journal of Endocrinology Prize Lecture. Eur J Endocrinol *155*, 645-653.

Sariola, H., and Saarma, M. (2003). Novel functions and signalling pathways for GDNF. J Cell Sci *116*, 3855-3862.

Sasselli, V., Pachnis, V., and Burns, A.J. (2012). The enteric nervous system. Dev Biol *366*, 64-73.

Schlee, S., Carmillo, P., and Whitty, A. (2006). Quantitative analysis of the activation mechanism of the multicomponent growth-factor receptor Ret. Nat Chem Biol *2*, 636-644.

Schlessinger, J., Plotnikov, A.N., Ibrahimi, O.A., Eliseenkova, A.V., Yeh, B.K., Yayon, A., Linhardt, R.J., and Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol Cell *6*, 743-750.

Schneider, R. (1992). The human protooncogene ret: a communicative cadherin? Trends Biochem Sci *17*, 468-469.

Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res *38*, W540-544.

Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F., and Pachnis, V. (1994). Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature *367*, 380-383.

Schuringa, J.J., Wojtachnio, K., Hagens, W., Vellenga, E., Buys, C.H., Hofstra, R., and Kruijer, W. (2001). MEN2A-RET-induced cellular transformation by activation of STAT3. Oncogene *20*, 5350-5358.

Scott, R.P., and Ibanez, C.F. (2001). Determinants of ligand binding specificity in the glial cell line-derived neurotrophic factor family receptor alpha S. J Biol Chem *276*, 1450-1458.

Seeliger, M.A., Young, M., Henderson, M.N., Pellicena, P., King, D.S., Falick, A.M., and Kuriyan, J. (2005). High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci *14*, 3135-3139.

Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., Grubel, G., Legrand, J.F., Als Nielsen, J., Colman, D.R., and Hendrickson, W.A. (1995). Structural basis of cell-cell adhesion by cadherins. Nature *374*, 327-337.

Shepherd, I.T., Beattie, C.E., and Raible, D.W. (2001). Functional analysis of zebrafish GDNF. Dev Biol *231*, 420-435.

Shepherd, I.T., Pietsch, J., Elworthy, S., Kelsh, R.N., and Raible, D.W. (2004). Roles for GFR{alpha}1 receptors in zebrafish enteric nervous system development. Development *131*, 241-249.

Shiu, S.H., and Li, W.H. (2004). Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. Mol Biol Evol *21*, 828-840.

Silvian, L., Jin, P., Carmillo, P., Boriack-Sjodin, P.A., Pelletier, C., Rushe, M., Gong, B., Sah, D., Pepinsky, B., and Rossomando, A. (2006). Artemin crystal structure reveals insights into heparan sulfate binding. Biochemistry *45*, 6801-6812.

Skinner, M.A., Safford, S.D., Reeves, J.G., Jackson, M.E., and Freemerman, A.J. (2008). Renal Aplasia in Humans Is Associated with RET Mutations. Am J Hum Genet *82*, 344-351.

Smith, D.P., Houghton, C., and Ponder, B.A. (1997). Germline mutation of RET codon 883 in two cases of de novo MEN 2B. Oncogene *15*, 1213-1217.

So, M.T., Leon, T.Y., Cheng, G., Tang, C.S., Miao, X.P., Cornes, B.K., Diem, N.N., Cui, L., Ngan, E.S., Lui, V.C.*, et al.* (2011). RET mutational spectrum in Hirschsprung disease: evaluation of 601 Chinese patients. PLoS ONE *6*, e28986.

Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem *287*, 252-260.

Stanley, P., Sundaram, S., and Sallustio, S. (1991). A subclass of cell surface carbohydrates revealed by a CHO mutant with two glycosylation mutations. Glycobiology *1*, 307-314.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography *25*, 495-503.

Takahashi, M., and Cooper, G.M. (1987). ret transforming gene encodes a fusion protein homologous to tyrosine kinases. Mol Cell Biol *7*, 1378-1385.

Takahashi, M., Ritz, J., and Cooper, G.M. (1985). Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell *42*, 581-588.

Takeuchi, K., Soda, M., Togashi, Y., Suzuki, R., Sakata, S., Hatano, S., Asaka, R., Hamanaka, W., Ninomiya, H., Uehara, H.*, et al.* (2012). RET, ROS1 and ALK fusions in lung cancer. Nat Med *18*, 378-381.

Tessier, D.C., Thomas, D.Y., Khouri, H.E., Laliberte, F., and Vernet, T. (1991). Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide. Gene *98*, 177-183.

Thornton, K., Kim, G., Maher, V.E., Chattopadhyay, S., Tang, S., Moon, Y.J., Song, P., Marathe, A., Balakrishnan, S., Zhu, H.*, et al.* (2012). Vandetanib for the treatment of symptomatic or progressive medullary thyroid cancer in patients with unresectable locally advanced or metastatic disease: U.S. Food and Drug Administration drug approval summary. Clin Cancer Res *18*, 3722-3730.

Treanor, J.J., Goodman, L., de Sauvage, F., Stone, D.M., Poulsen, K.T., Beck, C.D., Gray, C., Armanini, M.P., Pollock, R.A., Hefti, F.*, et al.* (1996). Characterization of a multicomponent receptor for GDNF. Nature *382*, 80-83.

Trupp, M., Arenas, E., Fainzilber, M., Nilsson, A.-S., Sieber, B.-A., Grigoriou, M., Kilkenny, C., Salazar-Grueso, E., Pachnis, V., Arumae, U.*, et al.* (1996). Functional receptor for GDNF encoded by the c-ret proto-oncogene. Nature *381*, 785-789.

Tsui-Pierchala, B.A., Ahrens, R.C., Crowder, R.J., Milbrandt, J., and Johnson, E.M., Jr. (2002). The long and short isoforms of Ret function as independent signaling complexes. J Biol Chem *277*, 34618-34625.

Tsuzuki, T., Takahashi, M., Asai, N., Iwashita, T., Matsuyama, M., and Asai, J. (1995). Spatial and temporal expression of the ret proto-oncogene product in embryonic, infant and adult rat tissues. Oncogene *10*, 191-198.

Ullrich, A., and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. Cell *61*, 203-212.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography *30*, 1022-1025.

van Heel, M. (1984). Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). Ultramicroscopy *13*, 165-183.

van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111-123.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M.*, et al.* (2000). Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys *33*, 307-369.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J Struct Biol *116*, 17-24.

van Weering, D.H., Moen, T.C., Braakman, I., Baas, P.D., and Bos, J.L. (1998). Expression of the receptor tyrosine kinase Ret on the plasma membrane is dependent on calcium. J Biol Chem *273*, 12077-12081.

Veiga-Fernandes, H., Coles, M.C., Foster, K.E., Patel, A., Williams, A., Natarajan, D., Barlow, A., Pachnis, V., and Kioussis, D. (2007). Tyrosine kinase receptor RET is a key regulator of Peyer/'s Patch organogenesis. Nature *446*, 547-551.

Virtanen, H., Yang, J., Bespalov, M.M., Hiltunen, J.O., Leppanen, V.M., Kalkkinen, N., Goldman, A., Saarma, M., and Runeberg-Roos, P. (2005). The first cysteine-rich domain of the receptor GFRalpha1 stabilizes the binding of GDNF. Biochem J *387*, 817-824.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860-864.

Wang, X., Baloh, R.H., Milbrandt, J., and Garcia, K.C. (2006). Structure of artemin complexed with its receptor GFRalpha3: convergent recognition of glial cell line-derived neurotrophic factors. Structure *14*, 1083-1092.

Webb, S.E.D., Roberts, S.K., Needham, S.R., Tynan, C.J., Rolfe, D.J., Winn, M.D., Clarke, D.T., Barraclough, R., and Martin-Fernandez, M.L. (2008). Single-Molecule Imaging and Fluorescence Lifetime Imaging Microscopy Show Different Structures for High- and Low-Affinity Epidermal Growth Factor Receptors in A431 Cells. Biophysical Journal *94*, 803-819.

Wehrman, T., He, X., Raab, B., Dukipatti, A., Blau, H., and Garcia, K.C. (2007). Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. Neuron *53*, 25-38.

Whitty, A., and Riera, T.V. (2008). New ways to target old receptors. Curr Opin Chem Biol *12*, 427-433.

Wiesner, S., Wybenga-Groot, L.E., Warner, N., Lin, H., Pawson, T., Forman-Kay, J.D., and Sicheri, F. (2006). A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. EMBO J *25*, 4686-4696.

Woods Group (2005-2012). GLYCAM WEB. Complex Carbohydrate Research Center, University of Georgia, Athens, GA (http://www.glycam.com).

Wybenga-Groot, L.E., Baskin, B., Ong, S.H., Tong, J., Pawson, T., and Sicheri, F. (2001). Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. Cell *106*, 745-757.

Yang, J., Runeberg-Roos, P., Leppanen, V.M., and Saarma, M. (2007). The mouse soluble GFR[alpha]4 receptor activates RET independently of its ligand persephin. Oncogene *26*, 3892-3898.

Yuzawa, S., Opatowsky, Y., Zhang, Z., Mandiyan, V., Lax, I., and Schlessinger, J. (2007). Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor. Cell *130*, 323-334.

Zhang, B., Luo, S., Wang, Q., Suzuki, T., Xiong, W., and Mei, L. (2008). LRP4 serves as a coreceptor of agrin. Neuron *60*, 285-297.

Zhang, J., Adrian, F., Jahnke, W., Cowan Jacob, S., Li, A., Iacob, R., Sim, T., Powers, J., Dierks, C., Sun, F.*, et al.* (2010). Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. Nature *463*, 501-506.

Zhang, X., Gureasko, J., Shen, K., Cole, P.A., and Kuriyan, J. (2006). An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell *125*, 1137-1149.

Zong, Y., Zhang, B., Gu, S., Lee, K., Zhou, J., Yao, G., Figueiredo, D., Perry, K., Mei, L., and Jin, R. (2012). Structural basis of agrin-LRP4-MuSK signaling. Genes Dev *26*, 247-258.